

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 30, 2004, 14:30:23 ; Search time 10660.7 Seconds  
(without alignments)  
5378.881 Million cell updates/sec

Title: US-09-744-167-2  
Perfect score: 6990  
Sequence: 1 MWIDENAVEDQLIKRYSW.....QLSEGPVVMELIFYILENIV 1323

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Dleop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_epool\_p/US09744167/runat\_29042004\_092919\_23693/app\_query.fasta\_1.3214  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744167.cgn\_1\_16435@runat\_29042004\_092919\_23693 -NCFU=6 -ICPU=3  
-NO MAP -LARGESQRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hug.\*  
3: gb\_in.\*  
4: gb\_cm.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hug\_hum.\*  
31: em\_hug\_inv.\*  
32: em\_hug\_other.\*  
33: em\_hug\_mus.\*  
34: em\_hug\_pln.\*  
35: em\_hug\_rtd.\*  
36: em\_hug\_nam.\*  
37: em\_hug\_vrt.\*  
38: em\_sv.\*  
39: em\_higo\_hum.\*  
40: em\_higo\_mus.\*  
41: em\_higo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6937	99.2	4769	9	BC032680 Homo sapi
2	6937	99.2	4839	9	AF104304 Homo sapi
3	6551	93.7	4565	9	AF130419 Homo sapi
4	3978.5	56.9	4823	5	AF104305 Xenopus l
5	3497	50.0	2625	9	AF130420 Homo sapi
6	3397	48.6	1977	6	AX809313 Sequence
7	3333	47.7	133799	9	AC105754 Homo sapi
8	3333	47.7	150129	2	AC026909 Homo sapi
9	2623	37.5	89973	9	AL591398 Human DNA
10	2471.5	35.4	85802	10	AL671309 Mouse DNA
11	2368.5	33.9	258787	2	AC095988 Rattus no
12	2056	29.4	201330	2	AC118106 Rattus no
13	2047	29.3	5480	10	BC042669 Mus muscu
14	2045	29.3	5438	10	AX122241 Mus muscu
15	2042	29.2	6280	9	BSM805726 Homo sapi
16	2039	29.2	4620	9	AF434817 Homo sapi
17	2039	29.2	6832	9	AB002303 Human mRN
18	2020.5	28.9	5841	9	HSN804398 Homo sapi
19	1620.5	23.2	3793	9	AK127003 Homo sapi
20	1597.5	22.9	4598	3	AF239997 Drosophil
21	1597.5	22.9	4551	3	DME310804 Drosophil
22	1585.5	22.7	4560	3	AY051821 Drosophil
23	1528.5	21.9	888	6	AX677329 Sequence
24	1466.5	21.0	67833	3	AC004564 Drosophil
25	1466.5	21.0	179139	3	AC099307 Drosophil
26	1466.5	21.0	313634	3	AE003454 Drosophil
27	1462	20.9	54186	2	AC019884 Drosophil
28	1312	18.8	1237	10	BC046341 Mus muscu
29	1172	16.8	887	10	BC058416 Mus muscu
30	1122.5	16.1	4499	9	BC032227 Homo sapi
31	1046	15.0	1144	10	BC038255 Mus muscu
32	998	14.3	576	6	AX677675 Sequence
33	945	13.5	192533	9	AC055876 Homo sapi
34	941	13.5	151800	9	AC138649 Homo sapi
35	940	13.4	178257	2	AC024930 Homo sapi
36	940	13.4	180741	9	AC091565 Homo sapi
37	930.5	13.3	228470	2	AC136776 Homo sapi
38	816.5	11.7	5158	9	HSN605324 Homo sapi
39	812.5	11.6	1915	3	AK114777 Clona int
40	809	11.6	164706	9	AC068614 Homo sapi
41	569	8.1	4275	3	AF045926 Caenorhab
42	521	7.5	175593	2	BS111079 Danio rer
43	469.5	6.7	193289	2	AC127065 Rattus no
44	427.5	6.1	139101	2	BS111174 Danio rer
45	403.5	5.8	2902	9	BC030808 Homo sapi

## ALIGNMENTS

RESULT 1

BC032680 4769 bp mRNA linear PRI 08-OCT-2003  
LOCUS Homo sapiens MAD, mothers against decapentaplegic homolog  
DEFINITION (Drosophila) interacting protein, receptor activation anchor, mRNA  
(cDNA clone MGC:45259 IMAGE:5585817), complete cds.  
ACCESSION BC032680.1 GI:21618607  
VERSION MGC.  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS  
1 (bases 1 to 4769)  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 4769)  
Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, R.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schaefer, R.E., Brownstein, M.J., Usdin, T.B., Tothiyuk, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalins, D.E.,  
Scherer, A., Schein, J.E., Jones, S.J., Skalska, U., Marra, M.A.,  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16906 (2002)  
22388257  
12477932  
2 (bases 1 to 4769)  
Strausberg, R.  
Direct Submission  
Submitted (06-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nri.nih.gov](mailto:nisc_mgc@nri.nih.gov)  
Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laxic, P., Legaspi, R.,  
Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, D.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 69 Row: p Column: 5  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4759059.  
Location/Qualifiers  
1..4769  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

BC032680 4769 bp mRNA linear PRI 08-OCT-2003  
LOCUS Homo sapiens MAD, mothers against decapentaplegic homolog  
DEFINITION (Drosophila) interacting protein, receptor activation anchor, mRNA  
(cDNA clone MGC:45259 IMAGE:5585817), complete cds.  
ACCESSION BC032680.1 GI:21618607  
VERSION MGC.  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS  
1 (bases 1 to 4769)  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 4769)  
Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, R.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schaefer, R.E., Brownstein, M.J., Usdin, T.B., Tothiyuk, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalins, D.E.,  
Scherer, A., Schein, J.E., Jones, S.J., Skalska, U., Marra, M.A.,  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16906 (2002)  
22388257  
12477932  
2 (bases 1 to 4769)  
Strausberg, R.  
Direct Submission  
Submitted (06-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nri.nih.gov](mailto:nisc_mgc@nri.nih.gov)  
Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laxic, P., Legaspi, R.,  
Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, D.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 69 Row: p Column: 5  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4759059.  
Location/Qualifiers  
1..4769  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

clone="MGC:45259 IMAGE:5585817"  
tissue\_type="Testis", embryonal carcinoma  
clone\_lib="NIH\_MGC\_92"  
lab\_host="DH10B"  
note="Vector: pCMV-SPORT6"  
1..4769  
/gene="MADHIP"  
note="synonyms: SARA, ZFYVE9"  
/db\_xref="LocusID:9372"  
/db\_xref="MIM:603755"  
12..4289  
/codon\_start=1  
/product="MADHIP protein"  
/protein\_id="AAH32680.1"  
/db\_xref="GI:21618608"  
/translation="MENYFQAEAYNLKVLDFEQNEDETVSTILLDTKNKILDPFS  
HRLSFTPLASVNESAVNESQPOLKVFSLAHSAPLTFEEDHCANODCNLNPFIAT  
XWIDENAVEDQLIKRYSWDDQCSAVEGEKGCNGLACLPDEKNVLVAVVHNCIDR  
TLONDLODNNYNSQSLMDAFSCPSQKDDGSGISGRDPSMAITSLTVDSVSSQGT  
PKTEGYSVNIPTSSSDSLASVCSQSLKDDGSGISGRDPSMAITSLTVDSVSSQGT  
GCPAYKAGNITPDEDLTGKISSPRTDLGSPNSHMGSEGLMKKEPAESTTESLR  
SGLPULLKPMFNGSRNDCRSCDLVFNVEVADNEGHEHEETIGTTEFLNMTX  
FSESQDMTNWKLTKLNENNDQVNEKEFLQISQPEDTNGSGCCVGLADAGLKH  
GTCTSESECDFTVIDTPAANYLNGCDYSQMDPGVFPKTLPSKESVLSQSSPKV  
EESSECYNIQEGNEATEGSLILNLTGDMKYNLHNFCSQVPSVLSQSSPKV  
ASLPSISVPGFARPKPSNLKQIPKPLSDHLONDLPANSNNNTKNDLRAQFGIGA  
ENSNATVCSPLSGNISVDTNGHELESYEAEISTRPCALADGSPDFCASCOSLACKLL  
RKPFITLGEVAPVWVDSQAQCMKCEARFTTKRSHCRACGVFCASCOSLACKLL  
YNDREARVVICHSVLMNAQAWENYMSASSQSPNNPNNPBYCSTIIPLOQAQASGL  
SSPPTVMVPGVLKHPGAQVREQRWFMADGILPNGEVADAAKLTMMGTSGAGT  
LAVSDHPKVTTSPLPAETDICTFQSVGSANLLIPEDGLPILISTGVK  
GDYAVEKPSQISVMQQLDGLPFLVFLNALLSMVKIYVYVNRKWCFTTGKMA  
VQSPBIVTLOCLPDKCLPKDIFNHFVQLYDALAGNVSNLGHVSRFSSFLGSKCH  
GGFLVYTSQSLQDLVLPYLLGILIQWETPAKVPFIRLMRLGAGRYLPYCP  
LFSVFRPLRFGEGHTIMLLADPRNYQVTLPVVQGLVDMVEYKTSIKTPSNRYNE  
MKKNNKSHVLAGGACFNEKADSHLVQVQNDGNGVQTOAISLHNPRTKVTGASFPV  
FSGALKSSGSLKSSIVEDGMVQITAEENDSLRQALREMKDFTTCGKADAEPOE  
HIHIQWDDKNSKGVSPIDGKSMETITNVKIFHGEYKANGKVRITWTEVFLEND  
DOHNCPLDPADSHRLTEHVAKAFCLALCPHLKLLKEDGMTKLGLRVLTDSDQVGYQAG  
SNGQPLPSQYMNLDLSALVPVHIGACQLSGPVMVMEILFYLEINIV"  
2091..2282  
/note="FYVE, Region: FYVE zinc finger. The FYVE zinc  
finger is named after four proteins that it has been found  
in: Fabi, YOTB/ZK632.12, Vac1, and EAL. The FYVE finger  
has been shown to bind two Zn++ ions. The FYVE finger has  
eight potential zinc coordinating cysteine positions. Many  
members of this family also include two histidines in a  
motif R-HHC-XCG, where + represents a charged residue and  
X any residue. We have included members which do not  
conserve these histidine residues but are clearly related"  
/db\_xref="CDD:pfam01363"

Alignment Scores:  
Pred. No.: 0 Length: 4769  
Score: 6337.00 Matches: 1315  
Percent Similarity: 99.47% Conservative: 1  
Best Local Similarity: 99.40% Mismatches: 7  
Query Match: 99.24% Indels: 0  
DB: 9 Gaps: 0  
US-09-744-167-2 (1-1323) x BC032680 (1-4769)  
QY 1 MetTPIleAspGluAsnAlaValGluAlaGluAspGlnLeuLysArgAsnTySerTrp 20  
Db 318 ATGTGGAATGATGAATAATGCTGTTGAGAGACCACTTAATTAAGAGAACTATAGTGG 377  
QY 21 AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu 40  
Db 378 GATGATCAATGAGTGTGTTGAGTGGAGAGAGAAATGTGGAAACCTGGCTGTCTG 437  
QY 41 ProAspGluLysAsnValLeuValValAlaValMetHisAsnCysAspLysArgThrLeu 60

Db 438 CCAGATGAGAGAAATGTTCTTTGTTAGCCGTCAATGATCAATTAACATGATGATAAAGCAATTA 497  
Qy 61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80  
Db 498 CAATACCGATTACAGGATTGTAATTAATAATAGTCAATCCCTTATGGAGTCTTTAGC 557  
Qy 81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100  
Db 558 TGTTCACTGGATATGAACACAGACAAATGATCAATTTAGTTTATGATAAATGAGTCC 617  
Qy 101 ThrGluAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu 120  
Db 618 ACTGAAAAGATATGAATTCAGAGAAACAAATGGATCCATTTGAATGACGAAACAGAG 677  
Qy 121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140  
Db 678 GGGAGATCTGTAAACCATCTGTCTTACCTTCACTGATAGTCTAGCCAGTGTCTGTTC 737  
Qy 141 ProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160  
Db 738 CCTTCACAATTAAGAGGATGACGGAGATATAGGTAGAGACCCCTCCATGTCTGGCATTA 797  
Qy 161 SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys 180  
Db 798 AGTTTACGGTTGATTCAGTAATCTCATCCAGGAAACAGATGATGCTCTGTGTTAA 857  
Qy 181 LysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysIleSerSerProArgThr 200  
Db 858 AAGCAAGAGAACTATATACAGATGAGGACCTCACTGGCAAAATCAGCTCTCTAGGACA 917  
Qy 201 AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu 220  
Db 918 GATCTAGGAGTCCAAATTCCTTTCCCATGATGAGGGGATTTTGATGAAAAGAG 977  
Qy 221 ProAlaGluGluSerThrGluGluSerLeuArgSerGlyLeuProLeuLeuLys 240  
Db 978 CCAGCAGAGAGACCACTGAAGATCCCTCCGGTCTGGTTTACCTTTGCTTCTCATA 1037  
Qy 241 ProAspMetProAsnGlySerGlyArgAsnAspCysGluArgCysSerAspCysLeu 260  
Db 1038 CCAGACATGCTATATGGGTCTGAAGGAATTAATGACTGTGAACGGTGTTCAGATTGCC 1097  
Qy 261 ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly 280  
Db 1098 GTGCCATTAAGTTAGGCTGATGAATAATGAAGTTATGAATGAAGAAATCTTTGCC 1157  
Qy 281 ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTrp 300  
Db 1158 ACTACAGAAATCCTTAATATGACAGACATTTCTCTGAATCTCAGGACATGACTAATTGG 1217  
Qy 301 LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluGluLysPhe 320  
Db 1218 AAGTTGACTAACTAAATGAGATGAATGATAGCCAAAGTAAACGAAAGGAAAGTTT 1277  
Qy 321 LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlyCysValGlyLeu 340  
Db 1278 CTACAGATTAGTCAGCTCAGGACACTAATGGTGAATGAGGACAGTGTGTGATGG 1337  
Qy 341 AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe 360  
Db 1338 GCAGATGCAGTCTAGATTTTAAAGGAACCTTGCAATGATGATAAGTGAAGAAATGTGATT 1397  
Qy 361 SerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCysAspSerTyrGly 380  
Db 1398 TCCACTGTTATAGACACACAGGAGCAATATCTATCTAATGTTGTGATTCCTATGGA 1457  
Qy 381 MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal 400  
Db 1458 ATGCAAGACCCAGGTGTTCTTTTGTGTTCCAAAGACTTTTACCTCCCAAGAGATTTCAGTA 1517  
Qy 401 ThrGluGluLysGluIleGluGluSerLysSerGluCysTyrSerAsnIleTyrGluGln 420

Db 1518 ACAGAGAAAAAGAAATAGAGAAAGCAAGTCAAGATGCTCAAAATATTTATGAAACAG 1577  
Qy 421 ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuLeuAsnSerThrGlyAspLeuMet 440  
Db 1578 AGAGGAATAGAGCCACAGAGAGAGTGGACTACTTATTAACAGCAGCTGGTCACTAATG 1637  
Qy 441 LysLysAsnTyrLeuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer 460  
Db 1638 AAGAAAAATTTATTCATAAATTTCTGTAGTCAAGTTCATCAGTCTTTGGGCAATCTTCC 1697  
Qy 461 ProLysValValAlaSerLeuProSerIleSerValProPheGlyGlyAlaArgProLys 480  
Db 1698 CCCAAGGTAGTAGCAAGCTGCCATCTATCAGTGTCTCTTTGGTGGTGCAGACCCCAAG 1757  
Qy 481 GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp 500  
Db 1758 CAACCTTCTAACTTAACTTCAAAATTCAAAGCCATATCAGACCAATTTACAAAATGAC 1817  
Qy 501 PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys 520  
Db 1818 TTTCTCTGCAAAACAGTGGAAATTAATACTAAAAATAAAATGATATTCTTTGGGAAACAAA 1877  
Qy 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540  
Db 1878 TTAGGGGAAACTCAGCAACCAATGTATGCAGTCCATCTTTGGGAAACATCTCTAATGTC 1937  
Qy 541 AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu 560  
Db 1938 GATACAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTT 1997  
Qy 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580  
Db 1998 GATTAGCTCCAGATAGCCAGCAATATGATCTCAGAGCTGGTCAGTTGGAAATTTCTGCC 2057  
Qy 581 ArgLysProPheThrThrLeuGlyGluValAlaProValTyrValProAspSerGlnAla 600  
Db 2058 AGAAAGCCATTCACCACTCTGGTCTGGTGGTCCAGTATGGTACCGGATTTCTCAGGCT 2117  
Qy 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg 620  
Db 2118 CCAAAATTCATGAAATGTGAAGCCAGGTTTACATTCACCAAAAGAGGACATCAGTCAGA 2177  
Qy 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTyrMet 640  
Db 2178 GCATGTGGAGGTTTCTGTGCTTCTGTGTAGCTGAAATGTAAATGTTTATACATG 2237  
Qy 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAla 660  
Db 2238 GACAGAAAGGAAGCTAGAGTGTGTAAATCTGCCATTCAGTCTCAATGAATGCTCAAGCC 2297  
Qy 661 TyrGluAsnMetMetSerAlaSerSerGlnSerProAsnProAsnAsnProAlaGluTyr 680  
Db 2298 TGGGAGAACATGATGAGTGCCTCAAGCCAGAGCCCTAACCCCTAACCAATCTCTGCTGAATAC 2357  
Qy 681 CysSerThrIleProProLeuGlnGlnAlaGlnAlaSerGlyAlaLeuSerSerProPro 700  
Db 2358 TGTTCTACTATCCCTCCCTTGAGCAAGCTCAGGCTCAGGAGCTCTCAGCTCTCCACT 2417  
Qy 701 ProThrValMetValProValGlyValLeuLysHisProGlyAlaGluValAlaGlnPro 720  
Db 2418 CCCACTGTGATGGTACCTGTGGAGTTTAAAGCACCCCTGGAGCAGAAAGTGGCTCAGGCC 2477  
Qy 721 ArgGluGlnArgArgValTyrPheAlaAspGlyIleLeuProAsnGlyGluValAlaAsp 740  
Db 2478 AGAGAGCAGAGCGGAGTTTGGTTGCTGATGGGATCTTGCCCAATGGAGAGTTGCTGAT 2537  
Qy 741 AlaAlaLysLeuThrMetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAsp 760  
Db 2538 GCAGCAAAATTAACAATGAATGAACCTTCTCTGAGGAAACCCCTGGCTGTGTCAACAGC 2597  
Qy 761 ProValLysProValThrThrSerProLeuProAlaGluThrAspIleCysLeuPheSer 780  
Db 2598 CCAGTCAAGCCCACTAACTACCACTCTCTACAGCAGACAGGATATTGTTCTATTCTCT 2657

781 GlySerIleThrGlnValGlySerProValGlySerAlaMetAsnLeuIleProGluAsp 800  
 Db GGGAGTATAACTCAGGTTGGAAGTCTGTTGGAAGTGCATGAATCTTATTCCTGAAGAT 2717  
 801 GlyLeuProPheIleLeuIleSerThrGlyValGlyGlyAspTyrAlaValGluGluLys 820  
 Db GGCCTTCCTCCATCTCCATCTCCATCTCCATCTCCATCTCCATCTCCATCTCCATCT 2777  
 821 ProSerGlnIleSerValMetGlnGlnLeuGluAspGlyGlyProAspProLeuValPhe 840  
 Db CCATCAGAGATTCAGTAATGCACAGCTTGAGGATGGTGGCCCTGACCCACCTGTATTT 2837  
 841 ValLeuAsnAlaAsnLeuLeuSerMetValIleValAsnTyrValAsnArgLysCys 860  
 Db GTTTTAAATGCAAAATTTGTTGTCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2897  
 861 TrpCysPheThrThrLysGlyMetHisAlaValGlyGlnSerGluIleValIleLeuLeu 880  
 Db TGGTGTTCACACCAAGGAGATGCAATGAGTGGTCACTCGAGATAGTCACTTCCTCA 2957  
 881 GlnCysLeuProAspGluLysCysLeuProLysAspIlePheAsnHisPheValGlnLeu 900  
 Db CAGTGTTCACCGATGAAAGTGTGTTGCCAAGGATATCTTTAATCACTTTGTGCAGCTT 3017  
 901 TyrArgAspAlaLeuAlaGlyAsnValValSerAsnLeuGlyHisSerPhePheSerGln 920  
 Db TATCGGATGCTCTGGCAGGAGATGCTGAGTGGTCACTTCGAGATAGTCACTTCCTCA 3077  
 921 SerPheLeuGlySerLysGluHisGlyGlyPheLeuTyrValThrSerThrTyrGlnSer 940  
 Db AGTTTCTTGGCAGTAAAGACATGGTGGATCTTATATGTGACATCTACCTACCACTCA 3137  
 941 LeuGlnAspLeuValLeuProThrProProTyrLeuPheGlyIleLeuIleGlnLysTrp 960  
 Db CTGCAAGACCTTAGTACCTCCCAACCCCACTTACTTGTGTTGGGATCTTATCCAGAAATGG 3197  
 961 GluThrProTyrAlaLysValPheProIleArgLeuMetLeuArgLeuGlyAlaGluTyr 980  
 Db GAACTCTTGGGCTAAGATATTTCTTATCCGTCTGATGTGAGACTTGGAGCTGAATAT 3257  
 981 ArgLeuTyrProCysProLeuPheSerValArgPheArgLysProLeuPheGlyGluThr 1000  
 Db CGACTTTATCCATGCCCACTATTTCAGTGTGAGATTTCCGAAGCCATGTTTGGAGAGAG 3317  
 1001 GlyHisThrIleMetAsnLeuAlaAspPheArgAsnTyrGlnTyrThrLeuProVal 1020  
 Db GGGCATACCATCATGAATCTTCTTGGAGACTTCAGAAATACCAATATACCTTGGCAGTA 3377  
 1021 ValGlnGlyLeuValValAspMetGluValArgLysThrSerLeuLysIleProSerAsn 1040  
 Db GTTCAAGGTTTGGTGGTGTATGATGAAAGTTCGGAAGTCTAGCATCAAAATTCACGCAAC 3437  
 1041 ArgTyrAsnGluMetMetLysAlaMetAsnLysSerAsnGluHisValLeuAlaGlyGly 1060  
 Db AGATCAATGATGATGAAGAGCCATGAACAAAGTCCAAATGAGCATGTCCTGGCAGAGGT 3497  
 1061 AlaCysPheAsnGluLysAlaAspSerHisLeuValCysValGlnAsnAspGlyAsn 1080  
 Db GCCTGCTTCAATGAAAGGCGAGACTCTCATCTTGTGTGTACAGAAATGATGATGAAAC 3557  
 1081 TyrGlnThrGlnAlaIleSerIleHisAsnGlnProArgLysValThrGlyValAsnPhe 1100  
 Db TATCAGACCCAGGATATCAGTATTCAATCAGCCAGAAAGATGATGTCCTGGCAGTTTC 3617  
 1101 PheValPheSerGlyAlaLeuLysSerSerSerGlyTyrLeuAlaLysSerIleVal 1120  
 Db TTTGTGTTTCACTGGCGCTCTGAATCTCTGATACCTTGGCAAGTCCAGTATTGTG 3677  
 1121 GluAspGlyValMetValGlnIleThrAlaGluAsnMetAspSerLeuArgGlnAlaLeu 1140  
 Db GAAGATGGTGTATGGTCCAGATTACTGCAGAAACATGGATTCTCTTGGAGCGAGCACTG 3737

1141 ArgGluMetLysAspPheThrIleThrCysGlyLysAlaAspAlaGluGluProGlnGlu 1160  
 Db CGAGAGATGAAGGACTTCCACCTGTGGAGGCGGAGCGGAGGAGAACCCAGAG 3797  
 1161 HisIleHisIleGlnTyrValAspAspLysAsnValSerLysGlyValValSerPro 1180  
 Db CACATCCACATCCAGTGGTGGATGATGACAAAGACGTTAGCAAGGCTGCTAAGTCT 3857  
 1181 IleAspGlyLysSerMetGluThrIleThrAsnValLysIlePheHisGlySerGluTyr 1200  
 Db ATAGATGGGAAGTCCATGAGACTATACAAATGTGAAGATATTCCATGATCAGAAATAT 3917  
 1201 LysAlaAsnGlyLysValIleArgTyrThrGluValPhePheLeuGluAsnAspGln 1220  
 Db AAGCAATGGAAGATTAATCAGATGGACAGAGTGTCTTTCTAGAAAACGATGACCA 3977  
 1221 HisAsnCysLeuSerAspProAlaAspHisSerArgLeuThrGluHisValAlaLysAla 1240  
 Db CACAATTCCTCAGTATCTCTGAGATCAGATCAGATGATGATGATGATGATGATGATG 4037  
 1241 PheCysLeuAlaLeuCysThrGlnLeuLysLeuLysGlyAspGlyMetThrLysLeu 1260  
 Db TTTTCCCTTGTCT 4097  
 1261 GlyLeuArgValThrLeuAspSerAspGlnValGlyTyrGlnAlaGlySerAsnGlyGln 1280  
 Db GGACTACCTGTGACACTTGTGACTCAGATCAGATCAGATCAGATCAGATCAGATCAGAT 4157  
 1281 HisLeuProSerGlnTyrMetAsnAspPheAspSerAspLeuValLysMetIleHisGly 1300  
 Db CCCCTTCCCTCAGTATGATGAGGCGCCGCTGCTGATGAGTATGATGATGATGATGAT 4217  
 1301 GlyAlaCysGlnLeuSerGluGlyProValValMetGluLeuIlePheTyrIleLeuGlu 1320  
 Db GGGGCTCCAGCTTGTGAGGCGCCGCTGCTGATGAGTATGATGATGATGATGATGATGAT 4277  
 1321 AsnIleVal 1323  
 Db AACATCGTA 4286

RESULT 2  
 AF104304 4839 bp mRNA linear PRI 03-JAN-1999  
 LOCUS Homo sapiens Smad anchor for receptor activation (SARA) mRNA,  
 complete cds.  
 ACCESSION AF104304  
 VERSION AF104304.1 GI:4092766  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 4839)  
 AUTHORS Tsukazaki,T., Chiang,T.A., Davison,A.F., Attisano,L. and Wrana,J.L.  
 TITLE SARA, a FIVE domain protein that recruits Smad2 to the TGFbeta  
 RECEPTOR  
 JOURNAL Cell 95 (6), 779-791 (1998)  
 MEDLINE 99081294  
 PUBMED 9865696  
 REFERENCE 2 (bases 1 to 4839)  
 AUTHORS Wrana,J.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-NOV-1998) Program in Developmental Biology, Hospital  
 for Sick Children, 555 University Avenue, Toronto, Ontario M6G 1X8,  
 Canada

FEATURES  
 source  
 1..4839  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /cell\_type="HepG2"  
 /gene="SARA"

gene

CDS	439...4410 /gene="SARA" /codon_start=1 /product="Smad anchor for receptor activation" /protein_id="AAC99462.1" /db_xref="GI:4092767" /translation="NMIDENVAEDDLIKRYSWDDCCSAVEGEKCGNLAFLPDEK NVLVAVMNCRLIQLNDQCNINQSMDAPSCSLDNNKQIQDFSFINSSTE KMNSEKQMDPLNRPKTEGVSVNLCTSDLSASVCSPLKDDSGIDGDFSMGSAIT SLTVDSVSIQGGDGPVAKQENYIPDEDLTKISIPRDLGSPNSFSHMEGLMK KEPAESTTIESLRSLPLLLKPDMPNGSRNDCSDCLVPNEVRADENEGYEHE ETLTGTEFLNMTBHFSESQDNTNWKLTCLNEMNDSONNEEKEFLQISQEPDINGDSG GOCGLADAGLDLKGTCISEBECDFSTVIDTPAANYLNSGCDSYGMQDGPVFPVKT LPKSDNVPTEKIEISKSECYNIYEQBNEATEGSLILNSTGDLMKVNLNFC QVPSVLGQSKVAVSLPSISVFFGAPKQPSNLKLIQPKPLSLHLQNDPFAANSNN TKNKDILGKAKIGENSATNVCPSLIGNISVDNTEGHELESYEAEISTRPLCALAPDS PDNDLGAQGIISARKEFTTLGEVAPVMVFDSPQPNCKMCEARFTTKRHHCRAQK VFCASCSLCKLLYMDRKEARVCVICHSLVMAQAWENMSSASSQSPNPNPAEYCS TIPLOOQASGALSPPPTVMVPGVLKHPGAEVAQPREORRWAFADGILPNGEVAD AAKLTWNGTSSAGTLAVSDHPVKPTTISPLAETDILCFSGSTOVQSVPGSAMNLI EGLPILISTGVKGVAVVEEKPSQISVMQOLEDGGPDLVFLVLANILLSMYKIVNYV NRKCMWFTTGKMAVQSEIVILLQCLPEKCPKDFIHFHFVQLYRDALAGNVSNLQ HSFSSQFLGSKHGFLYVTSYQSLQDLVLPYLFGLIQKWTETPAKVPFPIRL MLRGAERYLPCLFSVRERKELFGETHTIMLLADFNRYQYTLPPVQGLVPMVEV RKTSLKIPSNRYNEMKAMNNEHVLGAGCFNEKADSHLVCVQNDGNYQTQALSI HNGPRKVTGASFVFGALKSSGYLAKSSIVEDGVMOVITAEENMDSLOALRMKDF TITCGADAEPEOEHLIOWDDDDKNVSKGVSPIDGKSMETITNVAIPHGSEYKANG KVIRWEVFFLENDGNNCLSDPADHSRLTEHVAKAPCLALCPHLKLLKEDGNTKGL RVILSDQVQYQNGSNGQLPSPQYMNDLDSALVPVLIHGGACQLSEGGPVMMELIFYILE NIY"	Alignment Scores: Pred. No.: 0 Length: 4839 Score: 6937.00 Matches: 1315 Percent Similarity: 99.47% Conservative: 1 Best Local Similarity: 99.40% Mismatches: 7 Query Match: 99.24% Indels: 0 DB: 9 Gaps: 0
ORIGIN	US-09-744-167-2 (1-1323) x AF104304 (1-4839)	
Qy	1 MetTrpIleAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrSerTrp 20	
Db	439 ATGTGGATTGATGAAAATGCTGTGTGAGAGAGACGAGTAAATTAAGAGAACTATAGTTGG 498	
Qy	21 AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu 40	
Db	499 GATGATCAATCGAGTGTGTGAAGTGGAGAGAAATGTGGAACCTGGCTTGTCTG 558	
Qy	41 ProAspGluLysAsnValLeuValAlaValMethHisAsnCysAspLysArgThrLeu 60	
Db	559 CCAGATGAGAAGAAATGTTCTTTGTTGTAGCGCTCATGCATAACTGTGATAAAGGACATTA 618	
Qy	61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMethAsnAlaPheSer 80	
Db	619 CAATAACGATTACAGGATTGATAATTAATAATAGTCAATCCCTTATGGATGCTTTTAGC 678	
Qy	81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100	
Db	679 TGTTCACTGGATAATGAAACACAGACAACTGATCAATTTAGTTTAGTATAAATGAGTCC 738	
Qy	101 ThrGluLysAspMetAsnSerGluLysGlnMethAspProLeuAsnArgProLysThrGlu 120	
Db	739 ACTGAAAAAGATATGAATTCAGAGAAACAAATGGATCCATTGTAATAGCCGAAACAGAG 798	
Qy	121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140	
Db	799 GGGAGATCTGTTAACCACTGTGTCTCTACTTCACTGATAGTCTAGCCAGTGTCTGTTC 858	
Qy	141 ProSerGlnLeuLysAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160	
Db	859 CTTTCCAAATTAAAGGATGACGGAAGTATAGGTAGAGACCCCTCCATGTCGCGATTACA 918	

QY 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540  
Db 1999 TTAGGGAAACTCAGCAACCAATGATGAGTCCATCTTTGGAAACATCTCTAATGTC 2058  
QY 541 AspThrAsnGlyGluHisLeuGluSerTyGluAlaGluIleSerThrArgProCysLeu 560  
Db 2059 GATCAAAATGGGAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCCIT 2118  
QY 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580  
Db 2119 GCATTAGCTCAGATAGCCAGATATGATCTCAGAGCTGGTCACTTTTGAATTTCTGCC 2178  
QY 581 ArgIysProPheThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla 600  
Db 2179 AGAAAGCCATTACACCGCTGGGTGAGTGCTCCAGTATGGGTACCGGATCTCTCAGGCT 2238  
QY 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg 620  
Db 2239 CCAAAATTCATGAATGTGAAGCCAGGTTTACATTCACCAAAAGGAGGCTACTGCAGA 2298  
QY 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTyrMet 640  
Db 2299 GCATGTGGGAAGTTTCTGTGCTTCTGCTGTAGCCTGAAATGTAAACTGTTATACATG 2358  
QY 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAla 660  
Db 2359 GACAAAGGAAGCTAGAGTGTGTAACTTGCCATTCAGTCTGATGAATGCTCAAGCC 2418  
QY 661 TrpGluAsnMetMetSerAlaSerSerGlnSerProAsnProAsnProAlaGluTyr 680  
Db 2419 TGGGAGAACATGATGAGTGCCTCAAGCCAGAGCCCTTAACCTTAACATCTGCTGAATAC 2478  
QY 681 CysSerThrIleProProLeuGlnGlnAlaGlnAlaSerGlyAlaLeuSerSerProPro 700  
Db 2479 TGTTCTACTATCTCCCTGCTGACCAAGCTCAGCGCTCAGGAGCTCTGAGCTCTCCACCT 2538  
QY 701 ProThrValMetValProValGlyValLeuLysHisProGlyAlaGluValAlaGlnPro 720  
Db 2539 CCCACTGTATGTTACCTGTGGAGTTTAAAGCACCTCGAGCAGAAAGTGGCTCAGCCC 2598  
QY 721 ArgGluGlnArgArgValTrpPheAlaAspGlyIleLeuProAsnGlyGluValAlaAsp 740  
Db 2599 AGAGAGCAGAGCGGAGTTTGGTTTGTCTGATGGGATCTTGCCCAATGGAGAAGTGTCTGAT 2658  
QY 741 AlaAlaLysLeuThrMetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAsp 760  
Db 2659 GCAGCCAAATTAACAATGAATGGAACTTCTCTGACAGAACCTTGCTGTGTGCACAGAC 2718  
QY 761 ProValLysProValThrThrSerProLeuProAlaGluThrAspIleCysLeuPheSer 780  
Db 2719 CCAGTCAAGCCAGTAATACCACTACCAGTCTTACCAGCAGAGCGGATATTTGCTATTCTCT 2778  
QY 781 GlySerIleThrGlnValGlySerProValGlySerAlaMetAsnLeuIleProGluAsp 800  
Db 2779 GGGAGTAAATCTCAGGTTGGAAGTCTGTTGGAAGTGAATGAATCTATTCTCTGAAGAT 2838  
QY 801 GlyLeuProPheIleLeuIleSerThrGlyValLysGlyAspTyAlaValGluGluLys 820  
Db 2839 GGCCTTCTCCATCTCATCTCCACTGGTGTAAAGAGAGACTATGCTGTGGAAGAGAA 2898  
QY 821 ProSerGlnIleSerValMetGlnGlnLeuGluAspGlyGlyProAspProLeuValPhe 840  
Db 2899 CCATCAGAGATTTTCAGTAATGACAGCTGGAGATGGTGGCCCTGACCCACATGTATTT 2958  
QY 841 ValLeuAsnAlaAsnLeuSerMetValLysIleValAsnTyValAsnArgLysCys 860  
Db 2959 GTTTTAATGCAATTTGTTGTAATGGTAAATTTGTAATTTATGTGAACAGAGAGTGC 3018  
QY 861 TrpCysPheThrThrLysGlyMetHisAlaValGlyGlnSerGluIleValIleLeuLeu 880  
Db 3019 TGGTGTCTTCAACAAAGGAATCATGCACTGGGTGAGTGTGAGATGATGATCTCTCTTA 3078  
QY 881 GlnCysLeuProAspGluLysCysLeuProLysAspIlePheAsnHisPheValGlnLeu 900

Db 3079 CAGTGTCTTACCGGATGAAAAGTGTTCCAAAGGATATCTTTAATCACTTTGTGCACTT 3138  
QY 901 TyrArgAspAlaLeuAlaGlyAsnValValSerAsnLeuGlyHisSerPhePheSerGln 920  
Db 3139 TATCGGATGCTCTGGCAGGGAATGTGGTAGAACATTTGGGACATTCCTCTTCTCAGTCAA 3198  
QY 921 SerPheLeuGlySerLysGluHisGlyGlyPheLeuTyrValThrSerThrTyGlnSer 940  
Db 3199 AGTTTCTCTGGCAGTAAAGAAACATGGTGGATCTTATATGTGACATCTACCTACAGTCA 3258  
QY 941 LeuGlnAspLeuValLeuProThrProProTyLeuPheGlyIleLeuIleGlnLysTrp 960  
Db 3259 CTGCAAGACCTAGTACTCCCAACCCCACTTACTTGTGTGGATCTTATCCAGAAATGG 3318  
QY 961 GluThrProTrpAlaLysValPheProIleArgIleuMetLeuArgLeuGlyAlaGluTyr 980  
Db 3319 GAAACTCTCTGGGCTAAAGTATTTCTATCCGCTCTGATGTTGAGACTTGGAGCTGAATAT 3378  
QY 981 ArgLeuTyrProCysProLeuPheSerValArgPheArgLysProLeuPheGlyGluThr 1000  
Db 3379 CGACTTTATCCATGCCCTACTATTCTAGTGTCAATTTGGAAGCCATTTGTTGGAGAGACG 3438  
QY 1001 GlyHisThrIleMetAsnLeuLeuAlaAspPheArgAsnTyGlnTyThrLeuProVal 1020  
Db 3439 GGGCATACCATCATGAATCTTCTTGACACTTCAGAAATTTACCATGATATACCTTCCAGTA 3498  
QY 1021 ValGlnGlyLeuValValAspMetGluValArgLysThrSerIleLysIleProSerAsn 1040  
Db 3499 GTTCAAGTGTGGTGGTTGATATGGAAGTTCGGAACCTAGCAAAATTTCCAGCAAC 3558  
QY 1041 ArgTyAsnGluMetMetLysAlaMetAsnLysSerAsnGluHisValLeuAlaGly 1060  
Db 3559 AGATACATGAGATGATGAAGCCATGAACAGTCCAAATGAGCATGCTCTGCAGGAGGT 3618  
QY 1061 AlaCysPheAsnGluLysAlaAspSerHisLeuValCysValGlnAsnAspAspGlyAsn 1080  
Db 3619 GCTCTCTTCAATGAAAAGCCAGACTCTCATCTGTGTGTGTACAGAATGATGATGGAAC 3678  
QY 1081 TyrGlnThrGlnAlaIleSerIleHisAsnGlnProArgLysValThrGlyAlaSerPhe 1100  
Db 3679 TATCAGACCCAGCTATCAGTATTTCACATCAGCCAGAAAGTGAAGTGGTCCAGTTTC 3738  
QY 1101 PheValPheSerGlyAlaLeuLysSerSerGlyTyLeuAlaLysSerSerIleVal 1120  
Db 3739 TTTGTCTCAGTGGGCTCTGAAATCTCTTCTGATACCTTGCAGTCCAGTATTTGTG 3798  
QY 1121 GluAspGlyValMetValGlnIleThrAlaGluAsnMetAspSerLeuArgGlnAlaLeu 1140  
Db 3799 GAAGATGTGTGTATGTTCCAGATCTACAGAACATGGAATCTCTTGGCAGGAGCTG 3858  
QY 1141 ArgGluMetLysAspPheThrIleThrCysGlyLysAlaAspAlaGluGluProGlnGlu 1160  
Db 3859 CGAGAGATGAAGACTTCCCATCACCCTGCGAAGCGGAGCGGAGAAACCCAGAG 3918  
QY 1161 HisIleHisIleGlnTrpValAspAspLysAsnValSerLysGlyValValSerPro 1180  
Db 3919 CACATCCCATCCAGTGGGTGATGATCAAGAACCTTAGCAAGGGTGTCTGTAAGTCTCT 3978  
QY 1181 IleAspGlyLysSerMetGluThrIleThrAsnValLysIlePheHisGlySerGluTyr 1200  
Db 3979 ATAGATGGGAAGTCCATCGAGACTATACAATGTGAAGATATTCATGGATCAGAAATAT 4038  
QY 1201 LysAlaAsnGlyLysValIleArgTrpThrGluValPhePheLeuGluAsnAspGln 1220  
Db 4039 AAAGCAAAATGCAAAAGTAAATCAGATGGACAGAGGTGTTTTTCTAGAAAACGATGACAG 4098  
QY 1221 HisAsnCysLeuSerAspProAlaAspHisSerArgLeuThrGluHisValAlaLysAla 1240  
Db 4099 CCAATTTGCTCAGTGATCTCTGCAGATCAGATGATGATGAGCATGTTGCCAAGCT 4158  
QY 1241 PheCysLeuAlaLeuCysThrGlnLeuLysLeuLeuLysGlyMetThrLysLeu 1260





Db 1139 CCAGCAGGAGAGCACCACCTGAAGAAATCCCTCGGTCTGGTTTACCTTTGCTTCTCAA 1198  
Qy 241 ProAspMetProAsnGlySerGlyValAsnAspCysGluArgCysSerAspCysLeu 260  
Db 1199 CCAGACATGCTAATGGTCTGGAGGAAATGACTGTGAACGGTGTTCAGATTGCCIT 1258  
Qy 261 ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly 280  
Db 1259 GTGCCTAATGAAGTTAGGGCTGATGAAATCAAGGTTATGAACATGAAGAACTCTTGGC 1318  
Qy 281 ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTrp 300  
Db 1319 ACTACAGAAATTCCTTAATATGACAGACATTTCTGTAACTCAGGACATGACTAATGG 1378  
Qy 301 LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluGluLysGluLysPhe 320  
Db 1379 AAGTTGACTAACTAAATGAGATGATGACCAAGTAAACGAAAGAAAGAAAGTTT 1438  
Qy 321 LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlnCysValGlyLeu 340  
Db 1439 CTACAGATTAGTCAGCCTGAGGACACTAATGGTGATAGTGGAGACAGTGTGTGGATTG 1498  
Qy 341 AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe 360  
Db 1499 GCAGATCGAGTCTAGATTAAAGAGAACTTGCATTAGTAGTGAAGTGAAGATGTGATTTC 1558  
Qy 361 SerThrValIleAspThrProAlaAsnTyrLeuSerAsnGlyCysAspSerTyrGly 380  
Db 1559 TCCACTGTTATAGACACACACAGCAGCAAAATTTATCTAATGGTTGTGATTCTCTATGGA 1618  
Qy 381 MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal 400  
Db 1619 ATGACAGACCCAGGTGTTCTTTGTTCCAAAGACTTTACCTCCAAAGAGAGATTCACTA 1678  
Qy 401 ThrGluGluLysGluIleGluSerLysSerGluCysTyrSerAsnIleTyrGluGln 420  
Db 1679 ACAGAGAAAGAAATAGAGAAAGCAAGTCAAGATGCTACTCAAAATATTTATGAACAG 1738  
Qy 421 ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuAsnSerThrGlyAspLeuMet 440  
Db 1739 AGAGAAATAGGCCACAGAGGAGTGGACTCTTTTAAACAGCAGCTGTGTGACCTAATG 1798  
Qy 441 LysLysAsnTyrLeuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer 460  
Db 1799 AAGAAATATTTATACATAATTTCTGTAGTCAAGTTCATCAGTGTCTTGGCAATCTTCC 1858  
Qy 461 ProLysValValAlaSerLeuProSerIleSerValProPheGlyAlaArgProLys 480  
Db 1859 CCCAAGGTAGTAGCAAGCCTGCCATCTATCAGTGTCTCTTTGGTGTGCAAGACCAAG 1918  
Qy 481 GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp 500  
Db 1919 CAACCTCTAATCTTAACCTCAATTCGAAGCCATTATCAGACCAATTTACAAATGAC 1978  
Qy 501 PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys 520  
Db 1979 TTTCTCTCAACACAGTGGAAATATACTAAAAATAAAATGATATTTCTTGGGAAAGCAAAA 2038  
Qy 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540  
Db 2039 TTAGGGGAAACTCAGCAACCAATGTATGAGTCCATCTTTGGGAAACATCTCTAATGTC 2098  
Qy 541 AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu 560  
Db 2099 GATCAAAATGGGAACATTTAGAAAGTTATGAGGCTGAGATCTCCATCAGACCATGCGCTT 2158  
Qy 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580  
Db 2159 GCATTAGCTCAGATAGCCCAAGTAATGATCTCAGAGCTGTGATTTGGAATTTCTGCC 2218  
Qy 581 ArgLysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla 600  
Db 2219 AGAAAGCAATTCACCACTCTGGGTGAGGTGCTCCAGTATGGGTACCGGATTCTCAGGCT 2278

Qy 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgHisHisCysArg 620  
Db 2279 CCNAATTCGATGAATGTGAAGCCAGGTTTACATTCCAAAGAGGAGCATCACTGCAGA 2338  
Qy 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLysMet 640  
Db 2339 GCATGTGGAAAGGTTTCTGTGCTTCGCTGPHAGGCTGAATGTAACCTGTTATACATG 2398  
Qy 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAla 660  
Db 2399 GACAGAAAGGAGCTAGAGTGTGTAACTGTCCATTGCGCTTAATGAAT 2449  
Qy 661 TrpGluAsnMetSerAlaSerSerGlnSerProAsnProAsnAsnProAlaGluTyr 680  
Db 2449 2449  
Qy 681 CysSerThrIleProProLeuGlnGlnAlaGlnAlaSerGlyAlaLeuSerSerProPro 700  
Db 2449 2449  
Qy 701 ProThrValMetValProValGlyValLeuLysHisProGlyAlaGluValAlaGlnPro 720  
Db 2450 2450  
Qy 721 ArgGluGlnArgValTrpPheAlaAspGlyIleLeuProAsnGlyGluValAlaAsp 740  
Db 2462 AGAGACGACAGCGGAGTTGGTTTGTGATGGATCTTGCCCAATGAGAGATTGCTGAT 2521  
Qy 741 AlaAlaLysLeuThrMetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAsp 760  
Db 2522 GCAGCCAAATTAACAAATGAATGAACTTCTCTCAGGAACCCCTGGCTGTGTACACGAC 2581  
Qy 761 ProValLysProValThrThrSerProLeuProAlaGluThrAspIleCysLeuPheSer 780  
Db 2582 CCAGTCAAGCAGTAACCTACAGTCTCTACAGCAGAGAGATATTTGCTATTCTCT 2641  
Qy 781 GlySerIleThrGlnValGlySerProValGlySerAlaMetAsnLeuIleProGluAsp 800  
Db 2642 GGGAGTATAACTCAGTTGGAGTCTCTGTGAAGTGCATGAATCTTATTCTTGAAGAT 2701  
Qy 801 GlyLeuProProIleLeuIleSerThrGlyValLysGlyAspTyrAlaValGluGluLys 820  
Db 2702 GSCCTTCTCCCAATCTCATCTCCATGGTGTAAAGAGAGACTATGCTGTGAAGAGAAA 2761  
Qy 821 ProSerGlnIleSerValMetGlnGlnLeuGluAspGlyGlyProAspProLeuValPhe 840  
Db 2762 CCATCAGAGATTTCACTAATGCAGCAGTTGGAGGATGGTGGCCCTGACCCACTTGTATT 2821  
Qy 841 ValLeuAsnAlaAsnLeuLeuSerMetValLysIleValAsnTyrValAsnArgLysCys 860  
Db 2822 GTTTTAAATGCAAAATTTCTGTCAATGGTTAAATTTGTAATTTATGTGAACAGAGAGTGC 2881  
Qy 861 TrpCysPheThrThrLysGlyMetHisAlaValGlyGlnSerGluIleValIleLeuLeu 880  
Db 2882 TGGTGTTCACACCAAGGGAATGCATGCAGTGGTCACTCTGAGATAGTCACTTCTCTA 2941  
Qy 881 GlnCysLeuProAspGluLysCysLeuProLysAspIlePheAsnHisPheValGlnLeu 900  
Db 2942 CAGTGTTTACCGGATGAAAGTGTTCGCAAGGATATCTTTAATCACTTTGTGCGAGCTT 3001  
Qy 901 TyrArgAspAlaLeuAlaGlyAsnValValSerAsnLeuGlyHisSerPhePheSerGln 920  
Db 3002 TATCGGGATGCTCTGGCAGGGAATGTGTGAGCAACTGGGACATTCCTTCTTCAGTCAA 3061  
Qy 921 SerPheLeuGlySerLysGluHisGlyGlyPheLeuTyrValThrSerThrTyrGlnSer 940  
Db 3062 AGTTTCTTGGCAGTAAAGAACATGGTGGATTCTATATGTGACATCTACCTACCACTCA 3121  
Qy 941 LeuGlnAspLeuValLeuProThrProTyrLeuPheGlyIleLeuIleGlnLysTrp 960  
Db 3122 CTGCAAGACCTAGTACTCCCAACCCCACTTACTTGTGTGGATTCTTATCCAGAAATGG 3181



QY 961 GluThrProThrAlaLysValPheProIleArgLeuMetLeuArgLeuGlyAlaGluTyr 980  
 Db |||||  
 3182 GAACCTCTTGGCTAAAGTATTTCTTATCCGCTGATGTTGAGACTTGGAGCTGAATAT 3241  
 QY 981 ArgLeuTyrProCysProLeuPheSerValArgPheArgLysProLeuPheGlyGluThr 1000  
 Db |||||  
 3242 CGACTTATCCATGCCACCTATTCACTGTCAGATTCGGAAGCCATTGTTGGAGAGACG 3301  
 QY 1001 GlyHisThrIleMetLeuLeuAlaAspPheArgAsnTyrGlnTyrThrLeuProVal 1020  
 Db |||||  
 3302 GGGCATACCATCATGAATCTTCTTGCAGACTTCAGAAATPACCAGATATACCTTGCAGGA 3361  
 QY 1021 ValGlnGlyLeuValValAspMetGluValArgLysThrSerIleLysIleProSerAsn 1040  
 Db |||||  
 3362 GTTCAAGGTTTGGTGTGATATGGAAGTTCGAAAACTAGCATCAAAATTCACCAAC 3421  
 QY 1041 ArgTyrAsnGluMetMetLysAlaMetAsnLysSerAsnGluHisValLeuAlaGlyGly 1060  
 Db |||||  
 3422 AGATACATCAGATGATGAAGCCATGAAACATGCAATGCAATGATGCTTGGCAGAGGT 3481  
 QY 1061 AlaCysPheAsnGluLysAlaAspSerHisLeuValCysValGlnAsnAspAspGlyAsn 1080  
 Db |||||  
 3482 GCCTGCTCTCAATGAAGGCAGACTCTCATCTTGTGTGTACAGATGATGATGAAC 3541  
 QY 1081 TyrGlnThrGlnAlaIleSerIleHisAsnGlnProArgLysValThrGlyAlaSerPhe 1100  
 Db |||||  
 3542 TATCAGACCCAGGCTATCAGTATTCACATCAGCCAGAAAGTGAATGCTGTCGACATTC 3601  
 QY 1101 PheValPheSerGlyAlaLeuLysSerSerGlyTyrLeuAlaLysSerSerIleVal 1120  
 Db |||||  
 3602 TTTGTGTTAGTGGCGCTCGAATCTCTTCTGATACCTTCCCAAGTCCAGTATTGTG 3661  
 QY 1121 GluAspGlyValMetValGlnIleThrAlaGluAsnMetAspSerLeuArgGlnAlaLeu 1140  
 Db |||||  
 3662 GAAGATGCTGTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 3721  
 QY 1141 ArgGluMetLysAspPheThrIleThrCysGlyLysAlaAspAlaGluProGlnGlu 1160  
 Db |||||  
 3722 A--GAGATGAAGGACTTCCATCATCCTGCTGGAAGCGGACCGAGGAAACCCAGGAG 3779  
 QY 1161 HisIleHisIleGlnTrpValAspAspLysAsnValSerLysGlyValValSerPro 1180  
 Db |||||  
 3780 CACATCCACATCCAGTGGTGGATGATGATGATGATGATGATGATGATGATGATGAT 3839  
 QY 1181 IleAspGlyLysSerMetGluThrIleThrAsnValLysIlePheHisGlySerGluTyr 1200  
 Db |||||  
 3840 ATAGATGGGAAGTCCATGGAGACTATTAACAAATGTGAAGATATTCATGATGATGAT 3899  
 QY 1201 LysAlaAsnGlyLysValIleAtrTrpThrGluValPhePheLeuGluAsnAspAspGln 1220  
 Db |||||  
 3900 AAGCAATGGAAGATTAATCAGATGATGATGATGATGATGATGATGATGATGATGAT 3959  
 QY 1221 HisAsnCysLeuSerAspProAlaAspHisSerArgLeuThrGluHisValAlaLysAla 1240  
 Db |||||  
 3960 CACAATTCCTCAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4019  
 QY 1241 PheCysLeuAlaLeuCysThrGlnLeuLysLeuLysGlyAspGlyMetThrLysLeu 1260  
 Db |||||  
 4020 TTTTGCCTTCT 4079  
 QY 1261 GlyLeuArgValThrLeuAspSerAspGlnValGlyTyrGlnAlaGlySerAsnGlyGln 1280  
 Db |||||  
 4080 GACTACGTGTGACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4139  
 QY 1281 HisLeuProSerGlnTyrMetAsnAspPheAspSerAspLeuValLysMetIleHisGly 1300  
 Db |||||  
 4140 CCCCCTCCCTCGCAGTACATGATGATGATGATGATGATGATGATGATGATGATGATG 4199  
 QY 1301 GlyAlaCysGlnLeuSerGluGlyProValValMetGluLeuIlePheTyrIleLeuGlu 1320  
 Db |||||  
 4200 GGGGCTTCCAGCTTAGTGGAGGCGCCGCTGATGATGATGATGATGATGATGATGATG 4259  
 QY 1321 AsnIleVal 1323

Db 4260 AACATAGTA 4268

# RESULT 4

AF104305

LOCUS

DEFINITION

Xenopus laevis Smad anchor for receptor activation (SARA) mRNA, complete cds.

ACCESSION

AF104305

VERSION

AF104305.1 GI:4092768

KEYWORDS

SOURCE

ORGANISM

Xenopus laevis (African clawed frog)

Xenopus laevis

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

REFERENCE

1 (bases 1 to 4823)

AUTHORS

TSUKAZAKI, T., CHIANG, T. A., DAVISON, A. F., ATTISANO, L. and WRANA, J. L.

TITLE

SARA, a FIVE domain protein that recruits Smad2 to the TGFbeta receptor

JOURNAL

Cell 95 (6), 779-791 (1998)

PUBMED

99081294

REFERENCE

2 (bases 1 to 4823)

AUTHORS

WRANA, J. L.

TITLE

Direct Submission

JOURNAL

Submitted 04-NOV-1998

Program in Developmental Biology, Hospital

for Sick Children, 555 University Avenue, Toronto, Ontario M6G 1X8,

Canada

FEATURES

Location/Qualifiers

1..4823

/organism="Xenopus laevis"

/mol\_type="mRNA"

/db\_xref="taxon:8355"

1..4823

/gene="SARA"

316..4023

/gene="SARA"

/codon\_start=1

/product="Smad anchor for receptor activation"

/protein\_id="AAC99463.1"

/db\_xref="GI:4092769"

/translation="MPKMWIGDTTMAEDSLFNFTGPTSEIVCNPTVESQSLQALDDQS

VNIHNEKSVLLADGFCPCSSPKSIINFDCLTMDNEMPLHSOMSVDDNDKEVTVLSVLP

TLIQTSNVSTDPALINKGTKEPHRALKETTSLVILPEIKPSTCAALSFPNNKVPYSY

QLNNTDLLSVSPVACSEQQQKTSLSHBEKLFEGVSATESFSAATAETVLDNLAIR

SABFDIVYKNFSDSCVINGDLTKSCGLSQESNEKFCASKFEFEGVDVNLNACVA

YKEADLPEENGNTAPMSLYNGCDSPGNPAQNPKNLPSKEDSVTEKEIEESKS

EYTGVEQQRDDVTERRGLLNKADQMKNNLHSLCNQVPSMHGQTSPPKKGITVQS

LSVPYGARTKQPTHLKLIHPKLSMLOSLIPENAGCSSKYKNDMLNKNQGNLI

SESLREDAVSPTVDANGDFEYRGFSICLAVSPSPNDLLAGAGVPISKPFT

TLGEVAVVVPDSQAPNCKEARFTFKRRHRCACVKFCAACCSLCKLQYMDKK

EARVCVICHSLVMAQWENMLSAVSQSPNPNFAYCYSTIPPMQQAQASGALSFPF

TMVPVGLVHPGTEGSKQRRVWFADGLPNEGTEADSDANNTVAGTLTVSHTN

NSTSESNSTGFCGSIQVGSANMLIPEDGLPILISTGVKDYAVFERPSQSMVQ

OLEEGGPPLVFLNLANLAVKIYVYVRCWCFTTKGMHAGVQAEIVILQCLPDE

KCLPDLPSHVELVQBAIAGNVGNLGHLSFLOSFLGSKDHGGFLVYVQSLQDL

VLPAPLYGLIIOKMETPWAKVPIRLMLRLGAEYRLVPCLEPSVRVKPLRGFTGH

TIINLADFRYQVILPVVGLVDMVRKTSIKIPSNRYNEMKAMNKSHEHLAIG

ACFNQMAADSHLVGVNDGNTQTAISIHQPKVTGTSFVFSKALSSGSLAKSS

IVEDGVVQITAESMDALRLREMKDFTTCGRADAEESQEHVYVQVDEKKNPKG

VFSFDGKSMESVTKIFHGSYKASGKIIRWIEVFFLDNEEQSGLSADHSRLT

ENKAFCLALCPHLKLLKEDGMTRILGRVSLDSQVGYQAGSNQQLLPARYTNDLGG

ALVPVTHGTCQLSGSPVSMELIFILENIS"

## ORIGIN

Alignment Scores:

Pred. No.: 7,39e-255 Length: 4823

Score: 3978.50 Matches: 825

Percent Similarity: 71.9% Conservative: 141

Best Local Similarity: 61.4% Mismatches: 245

Query Match: 56.9% Indels: 131

DB: Gaps: 26

US-09-744-167-2 (1-1323) x AF104305 (1-4823)

Qy 1 MetTrpIleAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrSerTrp 20  
Db 325 ATGTGTATGAGACACCATATATGCGAGAGGATTCACCTTTTAAACAACACTGGTCCCTCT 384  
Qy 21 AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu 40  
Db 385 GAAATTGTATCAACCCATCTACTGAGAGTCAA-----AGTTTACAGCTTTA 435  
Qy 41 ProAspGluLysAsnValLeuValAlaValMethHisAsnCysAspLysArgThrLeu 60  
Db 436 GATGATCAATCA-----GTGAATATTCACATGAAAAAGGTGTTCTCCTC 480  
Qy 61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80  
Db 481 GCTGATGGCTTTTCCACCATGCAGT-----AGCCCCAAAAGTATTATAAAC-----TTTGAC 531  
Qy 81 Cys---SerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGlu 99  
Db 532 TGCCTGACCATGATACGAAATGCCCTTGCACAGTCAATGAGTGTGTGATGACATGAC 591  
Qy 100 SerThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThr 119  
Db 591 ----- 591  
Qy 120 GluGlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCys 139  
Db 592 -----AAGAAACTGTPAACAAATTCAGTC 615  
Qy 140 SerProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSer----- 157  
Db 616 CTTCACAAATCATACAGGATACGTAGTAACGTAGACACAGCACCCAGCTATCAATAAACCT 675  
Qy 158 -----AlaIleThrSerLeuThrValAspSerValIleSerSer 170  
Db 676 GGCACTAAGAACCCCATAGACATTAAGAAACACCATCAGTTATTCGCTGAAATA 735  
Qy 171 GlnGlyThrAspGlyCysProAlaValLys---LysGlnGluAsnTyrIleProAspGlu 189  
Db 736 AAGCCTTACTCCACATGTGCTGCCCTTCGTTTGAAATTAACATAAGGTATCCAGTTAT 795  
Qy 190 AspLeuThrGlyLysIleSerSerProArgThrAspLeuGlySerProAsnSerPheSer 209  
Db 796 CAATTAAT-----AATACAGATCTA-----CTCAGCGTTTCA 828  
Qy 210 HisMetSerGluGlyIleLeuMetLysLysGluProAlaGluGluSerThrThrGluGlu 229  
Db 829 CCAGTGGTTGAAGCATGTAGTACAGCAGCAAAACATACATCTCTCTTCGATGAAGAA 888  
Qy 230 SerLeuArgSerGlyLeuProLeuLeuLysProAspMetProAsnGlySerGlyArg 249  
Db 889 AAATTTTGAAGGTGT-----TCTGCAACG 915  
Qy 250 AsnAsnAspCysGluArgCysSerAspCysLeuValProAsnGluValArgAlaAspGlu 269  
Db 916 GAGTCTTTGAGCCACTGTGCGGAACTGTACTGGATAAT----- 957  
Qy 270 AsnGluGlyTyrGluHisGluGluThrLeuGlyThrThrGluPheLeuAsn---MetThr 288  
Db 958 -----GAGGCTCTCCGTAGTCTCAATCTTTTGACATGTTGTA 996  
Qy 289 GluHisPheSerGluSerGlnAspMetThrAsnTrpLysLeuThrLysLeuAsnGluMet 308  
Db 997 AAGAACTTTTCTGACTCT---TGTGTGATTAATGCGGACTTGACTAAAAGTTGTGCGCTC 1053  
Qy 309 AsnAspSerGlnValAsnGluLysGlyPheLeuGlnIleSerGlnProGluAsp 328  
Db 1054 TCT-----CAAGAAAGCAATGAAAGTTTGTGCAAGTAAAGATTGAA--- 1098  
Qy 329 ThrAsnGlyAspSerGlyGlyGlnCysValGlyLeuAlaAspAlaGlyLeuAspLeuLys 348  
Db ----- 2187

Db 1099 -----GGAGGGGTAGATGCTAATGTCTTGTGAA 1128  
Qy 349 GlyThrCysIleSerGluSerGluCysAspPheSerThrValIleAspThrProAla 368  
Db 1129 AATGCTATGTAGCTTATAAAGAACATAGATTGGCTGAAGAAAATGGAACTAATGCA 1188  
Qy 369 AlaAsnTyrLeuSerAsnGlyCysAspSerTyrGlyMetGlnAspProGlyValSerPhe 388  
Db 1189 CCAATGCTCTGTACAAATGGGTGATTCCTATGGAATGAAAAACCCAGCCGTAGCTCAA 1248  
Qy 389 ValProLysThrLeuProSerLysGluAspSerValThrGluGluLysGluIleGluGlu 408  
Db 1249 AACCCAAAGAAATTTACCTTCAAAAGAAATCTGTGACAGAGAAAAGAAATTTGAAGAA 1308  
Qy 409 SerLysSerGluCysTyrSerAsnIleTyrGlu---GlnArgGlyAsnGluAlaThrGlu 427  
Db 1309 AGCAAGTCAGAAATCTATCTGTTTATGAACAACAAGAGAGATGATTACAGAG 1368  
Qy 428 GlySerGlyLeuLeuLeuAsnSerThrGlyAspLeuMetLysLysAsnTyrLeuHisAsn 447  
Db 1369 AGAGGTGCACCTTCTGTTAAATGCTAAGGCTCACCAATGAAGAAACAAT---TTGCATAGT 1425  
Qy 448 PheCysSerGlnValProSerValLeuGlyGlnSerSerProLysValValAlaSerLeu 467  
Db 1426 CTTTGAATCATGGTTCCATCCATGGCAACATCACCAAAAAGGGCAAGATTGTG 1485  
Qy 468 ProSerIleSerValProPheGlyValAlaArgProLysGlnProSerAsnLeuLysLeu 487  
Db 1486 CAATCTCTCAGTTTCCATACGGTGGAGCACGCACTAAGCAGCACTCATCTCAAACTC 1545  
Qy 488 GlnIleProLysProLeuSerAspHisLeuGlnAsnAsp---PheProAlaAsnSerGly 506  
Db 1546 CATATTTCCAAAGCCATTTCTGAAATGTTGCAGAGCGATCTCATCTCTCCAAATGCTGGC 1605  
Qy 507 AsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLysLeuGlyGluAsnSerAla 526  
Db 1606 TGCAGCTCTAAATACAAAATGACATGTTAAACAATCAATCAGGGGGAT----- 1656  
Qy 527 ThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal-----Asp 541  
Db 1657 ---AACCTGATTTCAGAATCACTGCGTGGAGGATCTGAGTGGCAGCCCTGTACTGAT 1713  
Qy 542 ThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeuAla 561  
Db 1714 GCTAATGTGTATTTCCCTGGAGAAATACAGGGGACCTGGCAGCTTG-----TGCCTTGCA 1767  
Qy 562 LeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAlaArg 581  
Db 1768 GTGTCTCCAGACAGCCAGACACAGATCTGCTTCCGCGGAGTTTGGGTACCCATCTCT 1827  
Qy 582 LysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGluAlaPro 601  
Db 1828 AAGCCATTTACTCTTAGGGGAAGTGGCTCAGTCTGGGTGGCAGATTCCCAAGCAGCA 1887  
Qy 602 AsnCysMetLysCysGluAlaArgPheThrLysArgArgHisHisCysArgAla 621  
Db 1888 AACTGCATGAAGTCGAGGCCAGATTACATTACCAAAAGAGGCGCATCACTGCGGAGCT 1947  
Qy 622 CysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLysMetAsp 641  
Db 1948 TGTGAAAAGGTGTTCTGTGCTGCTTGTGTCAGTCTAAATGCAAACTACAGTACATGAT 2007  
Qy 642 ArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAlaTrp 661  
Db 2008 AAAAAGGAGGCTCGTGTGTGTGTTATTGTCATTCTGTCTTATGATGCTCAAGCATGG 2067  
Qy 662 GluAsnMetSerAlaSerSerGlnSerProAsnProAsnProAlaGluTyrCys 681  
Db 2068 GAGAACATGTTAAGTGCATCGTCCAAAGCCCAATCAAAATAATCTCTGCTGAATACTGC 2127  
Qy 682 SerThrIleProLeuGlnGlnAlaSerGlyAlaLeuSerSerProProPro 701  
Db 2128 TCACTATCTCCATGACAGGACACAGCTTCAGAGCACTGAGTTCCCACTCC 2187





Db	1799	AGAAAAATATTATTACATTAATTTCTGTAGTCAAGTTCCATCAGTGTGGCGAATCTTCC	1858
Qy	461	ProLysValValAlaSerLeuProSerIleSerValProPheGlyGlyAlaAaGProLys	480
Db	1859	CCCAAGGTAGTAGCAGACCTGCCATCTATCAGTGTTCCTTTGGTGGTGAAGCCCAAG	1918
Qy	481	GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp	500
Db	1919	CAACCTTCTAATCTTAAACTTCAAAITCCAAAGCCATATTACAGACCATTTACAAAATGAC	1978
Qy	501	PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys	520
Db	1979	TTTCTCTGCAACAGTGGAAATATATCTPAAPAAATAAATGATATTCTTGGGAAGCAAAA	2038
Qy	521	LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal	540
Db	2039	TTAGGGGAAAACCTCAGCAACCAATGTATGCAGTCCATCTTTGGGAAACATCTCTAATGTC	2098
Qy	541	AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu	560
Db	2099	GATACAAATGGGGAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTGACCATGCGTT	2158
Qy	561	AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla	580
Db	2159	GCATTAGCTCCAGATAGCCAGATAATGATCTCAGAGCTGGTCAGTTTGGAAATTTCTGCC	2218
Qy	581	ArgLysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla	600
Db	2219	AGAAGCCATTCACCACTCTGGGTGAGGTGGCTCCAGTATGGGTACCGGATTTCTAGGCT	2278
Qy	601	ProAsnCysMetLysCysGluAlaAaGpPheThrPheThrLysArgArgHisCysArg	620
Db	2279	CCAAATGTCATGAATGTGAAGCCAGGTTTACATTACCAAAAGGAGGCATCTGCAGA	2338
Qy	621	AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTyrMet	640
Db	2339	GCATGTGGGAAGGTTTCTGTGCTTCTGCTGCTAGCCTGAAATGAACTGTTATACATG	2398
Qy	641	AspArgLysGluAlaAaGValCysValIleCysHisSerValLeuMetAsnAlaGlnAla	660
Db	2399	GACAGAAGAAGCTAGAGTGTGTAAATCTGCCATTCAGTGTCTAATGAATGGTAGTAT	2458
Qy	661	TrpGluAsnMetMetSer 666	
Db	2459	TAAACACAGGTTGATTCA 2476	
RESULT 6			
AX809313		1977 bp DNA linear	PAT 25-NOV-2003
LOCUS			
DEFINITION		Sequence 10 from Patent WO03045990.	
ACCESSION		AX809313	
VERSION		AX809313.1	GI:38523628
KEYWORDS			
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		1	
JOURNAL		Legrain, P., Gauthier, J.M., Collard, F. and Jacq, X.	
FEATURES		Protein-protein interactions involving transforming growth factor	
source		beta signalling	
		Patent: WO 03045990-A 10 05-JUN-2003;	
		Hybrigenics (FR)	
		Location/Qualifiers	
		1..1977	
		/organism="Homo sapiens"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:9606"	
ORIGIN			
Alignment Scores:		1.1e-216	1977
Pred. No.:			

Db 961 CACATATTAGTGTGAGATTTGGAGAGCCATTTGTTGGAGAGCGGGCATACCATCATG 1020  
 QY 1006 AsnLeuLeuAlaAspPheArgAsnTyrGlnTyrThrLeuProValGlnGlnGlyLeuVal 1025  
 Db 1021 AATCTTCTTGCAGACTTCAGAAATACAGATATACCTTGCAGTAGTTCAGGTTCAGTG 1080  
 QY 1026 ValAspMetGluValArgGlyThrSerIleLysLeuProSerAsnArgTyrAsnGluMet 1045  
 Db 1081 GTTGATATGGAGTTTCGGAATCTAGCATCAAAATCCACAGACAGATCAATGAGATG 1140  
 QY 1046 MetLysAlaMetAsnLysSerAsnGluHisValLeuAlaGlyGlyAlaCysPheAsnGlu 1065  
 Db 1141 ATGAAGCATGAACCAAGTCCAAATGAGCATGCTCTGGCAGAGAGTGTCTGCTTCATGAA 1200  
 QY 1066 LysAlaAspSerHisLeuValCysValGlnAsnAspAspGlyAsnTyrGlnThrGlnAla 1085  
 Db 1201 AAGGCAGACTCTCATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260  
 QY 1086 IleSerIleHisAsnGlnProArgLysValThrGlyAlaSerPhePheValPheSerGly 1105  
 Db 1261 ATCAGTATTCAAATCAGCCAGAAAGTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320  
 QY 1106 AlaLeuLysSerSerGlyTyrLeuAlaLysSerSerIleValGluAspGlyValMet 1125  
 Db 1321 GCTCTGAATCTCTTCTGTGATACCTTGCCTGAGTCCAGTATTTGTGGAAGATGGTGTATG 1380  
 QY 1126 ValGlnIleThrAlaGluAsnMetAspSerLeuArgGlnAlaLeuArgGluMetLysAsp 1145  
 Db 1381 GTCCAGATTACTGCAGAGAAATGATGATTCCTTGTAGGAGGAGTCCGAGAGATGAAGAC 1440  
 QY 1146 PheThrIleThrCysGlyLeuAlaAspAlaGluProGlnGlnHisIleGln 1165  
 Db 1441 TTCACCATCACTGTGGAGAGCGGAGCGGAGAGAACCCAGAGGACATCCATCCAG 1500  
 QY 1166 TrpValAspAspLysAsnValSerLysGlyValValSerProLysGlyLysSer 1185  
 Db 1501 TGGTGTGATGATGACAGAACTGTAGCAAGGCTGTCTGATGATGATGATGATGATGATG 1560  
 QY 1186 MetGluThrIleThrAsnValLysIlePheHisGlySerGluTyrLysAlaAsnGlyLys 1205  
 Db 1561 ATGAGACTATACAAATGTGAGATATTCATGATCAGATATTAAGCAATATGGAATA 1620  
 QY 1206 ValIleArgTTPThrGluValPheLeuGluAsnAspAspGlnHisAsnCysLeuSer 1225  
 Db 1621 GTAATCAGATGCAGAGGTGTTTCTAGAAACGATGACACGACCAATTCCTCAGT 1680  
 QY 1226 AspProAlaAspHisSerArgLeuThrGluHisValAlaLysAlaPheCysLeuAlaLeu 1245  
 Db 1681 GATCCTGAGATCACAGTAGATTCAGTACGATGTTGCCAAGCTTTTTCCTGTGTCTC 1740  
 QY 1246 CysThrGlnLeuLysLeuLysGlyAspGlyMetThrLysLeuGlyLeuArgValThr 1265  
 Db 1741 TGTCTCACCTGAAACTTCAGAGAGAGATGGAATGGAATGGAATGGAATGGAATGGAATG 1800  
 QY 1266 LeuAspSerAspGlnValGlyTyrGlnAlaGlySerAsnGlnHisLeuProSerGln 1285  
 Db 1801 CTTCAGCTCAGATCAGTGGTGTATCAGAGAGGAGCAATGGCCGCTTCCTCCTGCAG 1860  
 QY 1286 TyrMetAsnAspPheAspSerAspLeuValLysMetIleHisGlyGlyAlaCysGlnLeu 1305  
 Db 1861 TACATGATGATCTGATAGGCTTGTGTGCGGTGATCATGAGAGGGGCTCCAGCTT 1920  
 QY 1306 SerGluGlyProValValMetGluLeuIlePheTyrIleLeuGluAsnIleVal 1323  
 Db 1921 AGTGAGGCGCCGTTGTGTCATGGAATCATCTTTTATATTTCTGGAACATCGTA 1974  
 RESULT 7  
 AC105754  
 LOCUS AC105754 133799 bp DNA linear PRI 18-DEC-2002  
 DEFINITION Homo sapiens chromosome 1 clone RP4-814E15, complete sequence.  
 ACCESSION AC105754 AL122009  
 VERSION AC105754.3 GI:27228871

HTG. Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 133799)  
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 133799)  
 AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JAN-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 REFERENCE 3 (bases 1 to 133799)  
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-FEB-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 REFERENCE 4 (bases 1 to 133799)  
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-DEC-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 COMMENT On Dec 18, 2002 this sequence version replaced gi:18921345.  
 ----- Genome Center  
 Center: University of Washington Genome Center  
 Center Code: UMG  
 Web site: http://www.genome.washington.edu  
 Contact: uwgctgs@u.washington.edu  
 Drafting Center: SC  
 ----- Project Information  
 Center project name: chr-1  
 Center clone name: RP4-814E15 (sc0216)  
 ----- Summary Statistics  
 Sequencing vector: plasmid; 567 of reads  
 Sequencing vector: plasmid; 108752; 447 of reads  
 Chemistry: Dye-terminator Big Dye; 50% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 133740 bases at least Q40  
 Consensus quality: 133797 bases at least Q30  
 Consensus quality: 133799 bases at least Q20  
 Insert size: 133799; sum-of-contigs  
 Quality coverage: 8.6x in Q20 bases; sum-of-contigs  
 -----  
 Overlapping Sequences:  
 5': RP4-800M22 AL139156, 42347-bp overlap  
 3': RP11-155018 AL513218, 2000-bp overlap  
 -----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.  
 -----  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.  
 -----

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI			BglII			HindIII		
SeqDerMap	EngPrnt	SeqDerMap	EngPrnt	SeqDerMap	EngPrnt	SeqDerMap	EngPrnt	SeqDerMap
5810	5855	7542	7542	1289	1229			
2184	2142	5671	5692	449	<800			
8065	8190	5077	5096	512	<800			
486	<800	76	<800	2814	2825			
106	<800	3560	3636	1247	1229			
4616	4708	414	<800	4695	4762			
14239	14077	1723	1693	3373	3356			
3925	3921	1216	1202	3274	3356			
983	984	2887	2916	3116	3125			
2594	2609	8059	7992	68	<800			
2871	2878	1854	1835	1886	1926			
2276	2254	4987	4925	3144	3125			
2711	2609	13600	12991	9189	9097			
4738	4708	2718	2784	1213	1229			
445	<800	2884	2916	1915	1926			
1531	1495	4547	4472	731	747			
3546	3538	467	<800	3148	3125			
3154	3144	900	919	3481	3356			
896	984	2356	2329	2570	2594			
802	884	134	<800	5185	5250			
11042	10836	4510	4472	868	865			
4907	4887	3070	3044	18109	18317			
6450	6524	2284	2329	12807	12493			
3079	3045	3751	3739	4972	4989			
532	<800	2586	2640	4206	4097			
773	777	1052	1063	374	<800			
1464	1495	1058	1063	1931	1926			
6249	6230	236	<800	11410	11255			

1821	1816	1089	1063	9089	9097
2545	2609	3536	3471	5486	5606
1540	1495	440	<800	92	<800
2491	2609	9271	9390	14507	14568
1578	1495	1482	1491	5171	5250
1510	1495	916	919	1939	1926
2118	2142	1271	1267	166	<800
1084	1065	340	<800	6161	6265
1697	1677	3217	3260		
3818	3823	2323	2329		
1156	1132	9366	9390		
10017	9930	159	<800		
4212	4165	3486	3471		
2425	2405	4158	4097		
2682	2609	1524	1491		
1017	984	9505	9390		
7981	8190	6646	6734		
421	<800	2639	2640		

## FEATURES

source	Location/Qualifiers
1..133799	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="1"
	/clone="RP4-814E15"
	/clone_lib="RPC1 human PAC library 4"
10029..10030	
/note="IS1 and 9-bp target site duplication excised from between bases 10029 and 10806"	
GGTAATGACTCCCACTTACTGATAGTGTTTATGTTCAGATAATCCCGATGACCTTG	
TCATGCAACTCCACCGATTTCAGAACACAGTAACTTCGGTCCGAGCTTCCCGAT	
GTTGCTCAGATTTCAGATTATGCTCATGCGCTGAGTGAAGCTTGCTGATAAC	

## misc\_feature

## Alignment Scores:

Pred. No.:	8.4e-210	Length:	133799
Score:	3333.00	Matches:	632
Percent Similarity:	96.95%	Conservative:	3
Best Local Similarity:	96.49%	Mismatches:	8
Query Match:	47.68%	Indels:	12
DB:	9	Gaps:	2

US-09-744-167-2 (1-1323) x AC105754 (1-133799)

QY	1	MetTTPileAspGluAsnAlaValAlaGluAspGlnLeuLeuLysArgAsnTyrSerTTP	20
Db	60325	ATGTGATTGATGAAATGCTGTTGCAGAACCCAGTTAATTAAGAGAACTATAGTTGG	60384
QY	21	AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu	40
Db	60385	GATGATCAATGGAGTCTGTTGAAGTGGAGAGAGAAATGGAAACCTGGCTTGCTG	60444
QY	41	ProAspGluLysAsnValLeuValAlaValMetHisAsnCysAspLysArgThrLeu	60
Db	60445	CCAGATGAGAGAGATGTTCTTTGTTAGCGGTCATGCAATGCTGATTAAGGACATTA	60504



QY 61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80  
 Db 60505 CAAACAGATTACAGAGTTGTAATTAATTAATAGTCAATCCCTATGGATGCTTTAGC 60564  
 QY 81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100  
 Db 60565 TGTTTCACTGGTAATTAACAAACAGACAAACTGATCAATTTAGTTAGTATAAATGAGTCC 60624  
 QY 101 ThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu 120  
 Db 60625 ACTGAAGAAGATATGAATTCAGAAACAATGATGATTCATTAAGACCCAAACAGAG 60684  
 QY 121 GlyArgSerValAsnHisLeuCyseProThrSerSerAspSerLeuAlaSerValCysSer 140  
 Db 60685 GGGAGATCTGTTAACCATCTGTCTACTTCACTCATCTAGTATGCTAGGCAGTGTCTGTTC 60744  
 QY 141 ProSerGlnLeuLysAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160  
 Db 60745 CTTTCACAAATTAAGAGATGACGGAAGTATAGTAGACCCCTCCATGCTCTGCGATTACA 60804  
 QY 161 SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys 180  
 Db 60805 AGTTTAACGGTTGATTCAGTAACTCTCATCCAGGGAACAGATGATGCTGCTGTCTAAA 60864  
 QY 181 LysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysIleSerSerProArgThr 200  
 Db 60865 AAGCAAGAGAATAATATACAGATGAGGACCTCACTGGCAAAATCAGCTCTCTTAGGACA 60924  
 QY 201 AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu 220  
 Db 60925 GATCTAGGGAGTCCAAATTCCTTTTCCACATGATGAGGGGATTTTGTATGAAACAGAG 60984  
 QY 221 ProAlaGluGluSerThrThrGluSerLeuArgSerGlyLeuProLeuLeuLys 240  
 Db 60985 CCAGCAGAGGAGAGCACCACCTGAAGAATCCCTCCGGTCTGGTTACCTTTGCTTCTCAA 61044  
 QY 241 ProAspMetProAsnGlySerGlyArgAsnAsnAspCysGluArgCysSerAspCysLeu 260  
 Db 61045 CCAGACATGCTTAATGGTCTGGAAGGAATAATGACTGTGAACGGTCTTTCAGATTCCTT 61104  
 QY 261 ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly 280  
 Db 61105 GTGCTTAATGAATGAGGCTGATGAAATGAAGGTTATGAACATGAAGAACTCTTGCC 61164  
 QY 281 ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTrp 300  
 Db 61165 ACTACAGAAATTCCTTAATATGACAGACATTTCTCGAATCTCAGGACATGACTAATGCG 61224  
 QY 301 LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluLysGluLysPhe 320  
 Db 61225 AAGTTGACTAACTAATAGATGAATGATGACCAAGTAAACGAGAAAGAAAGTTT 61284  
 QY 321 LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlnCysValGlyLeu 340  
 Db 61285 CTACAGATTAGTACGCTGAGGACACTAATGGTATAGTGGAGGACAGTGTGTGATTG 61344  
 QY 341 AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe 360  
 Db 61345 GCAGATGAGGCTAGATTAAAGGAACTTGATAGTGAAGTGAAGATGATGATTTC 61404  
 QY 361 SerThrValIleAspThrProAlaAsnTyrIleSerAsnGlyCysAspSerThrGly 380  
 Db 61405 TCCACTGTTATAGACACACAGCAGCAATATCTATCTAATGTTGTCATTCCTATGGA 61464  
 QY 381 MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal 400  
 Db 61465 ATGCAAGACCCAGGTGTTCTTTGTTTCCAAAGACTTTACCTCCAAAGAAAGATTGATGA 61524  
 QY 401 ThrGluGlyLysGluIleGluGluSerLysSerGluCysThrSerAsnIleThrGluGln 420  
 Db 61525 ACAGAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAATATTTATGAACAG 61584  
 QY 421 ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuAsnSerThrGlyAspLeuMet 440

Db 61585 AGAGGAAATGAGGCCACAGAGGGAGTGACTACTTTTAAACAGCAGCTGGTACCTAATG 61644  
 QY 441 LysLysAsnTyrLeuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer 460  
 Db 61645 AAGAAAAATTAATTCATTAATTTCTGTAGTCAAGTTCCATCAGTCTGGGCAATCTTCC 61704  
 QY 461 ProLysValValAlaSerLeuProSerIleSerValProPheGlyGlyAlaArgProLys 480  
 Db 61705 CCCAAGGTAGTAGCAGCCTGCCATCTATCATGTTCTTTTGGTGGTGAAGACCCCAAG 61764  
 QY 481 GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp 500  
 Db 61765 CAACCTTCTAATCTTAAACTTCAAATTCAAAGCCATATCAGACCAITTTACAAATGAC 61824  
 QY 501 PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys 520  
 Db 61825 TTCTCTGCAAAACAGTGGAAATAATACTAAATAATAAATGATATTTCTTGGGAAAGCAAAA 61884  
 QY 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540  
 Db 61885 TTAGGGGAAACTCAGCAACCAATGTATGAGTCCATCTTTGGGAAACATCTCTAATGTC 61944  
 QY 541 AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu 560  
 Db 61945 GATACAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTT 62004  
 QY 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580  
 Db 62005 GCATTAGTCTCCAGATAGCCAGATATGATCTCAGAGCTGTCAGTTTGGATTTCTGCC 62064  
 QY 581 ArgLysProPheThrThrLeuGlyGluValalaProValTrpValProAspSerGlnAla 600  
 Db 62065 AGAAAGCCATTCCACACTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 62124  
 QY 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg 620  
 Db 62125 CCAATTTGATGAAATGTGAAGCCAGGTTTACATTCACAAAGAGGAGCATCTGCAGA 62184  
 QY 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuMet 640  
 Db 62185 GCATCTGGGAAAGTGAAGTTGC-----ATGTATACA 62214  
 QY 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeu 655  
 Db 62215 CTCAGAAAT-----CGGGCATGCACATTTGTATGCTGAATTA 62253  
 RESULT 8  
 AC026909  
 LOCUS  
 DEFINITION  
 Homo sapiens chromosome 1 clone RP11-174F10 map 1, WORKING DRAFT  
 SEQUENCE, 27 unordered pieces.  
 ACCESSION  
 AC026909  
 VERSION  
 AC026909.2 Gi:7596849  
 KEYWORDS  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 150129)  
 AUTHORS  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
 Boguslavsky, L., Boukhgaiter, B., Brown, A., Burkett, G.,  
 Campione, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Iliev, I., Johnson, K., Jones, C., Kann, L., Karatas, A.,  
 2 (bases 1 to 150129)  
 REFERENCE  
 2 (bases 1 to 150129)  
 AUTHORS  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
 Boguslavsky, L., Boukhgaiter, B., Brown, A., Burkett, G.,  
 Campione, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Iliev, I., Johnson, K., Jones, C., Kann, L., Karatas, A.,

Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehotsky, J.,  
 Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 McElm, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodores, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
 Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 19, 2000 this sequence version replaced gi:7328772.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L7273  
 Center clone name: L74\_F10  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 134873 bases at least Q40  
 Consensus quality: 142261 bases at least Q30  
 Consensus quality: 145325 bases at least Q20  
 Insert size: 155000; agarose-fp  
 Insert size: 147529; sum-of-contigs  
 Quality coverage: 3.7 in Q20 bases; agarose-fp  
 Quality coverage: 3.9 in Q20 bases; sum-of-contigs

TITLE  
 JOURNAL  
 COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 27 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1572: contig of 1572 bp in length  
 1573 1672: gap of 100 bp  
 1673 2741: contig of 1069 bp in length  
 2742 2841: gap of 100 bp  
 2842 4870: contig of 2029 bp in length  
 4871 4970: gap of 100 bp  
 4971 6056: contig of 1086 bp in length  
 6057 6156: gap of 100 bp  
 6157 7419: contig of 1263 bp in length  
 7420 7519: gap of 100 bp  
 7520 9305: contig of 1786 bp in length  
 9306 9405: gap of 100 bp  
 9406 11199: contig of 1794 bp in length  
 11200 12299: gap of 100 bp  
 12300 13054: contig of 1755 bp in length  
 13055 13154: gap of 100 bp  
 13155 15817: contig of 2663 bp in length  
 15818 15917: gap of 100 bp  
 15918 18265: contig of 2348 bp in length  
 18266 18365: gap of 100 bp  
 18366 20797: contig of 2432 bp in length  
 20798 20897: gap of 100 bp  
 20898 24397: contig of 3500 bp in length  
 24398 24497: gap of 100 bp  
 24498 27417: contig of 2920 bp in length  
 27418 27517: gap of 100 bp

27518 29053: contig of 1536 bp in length  
 29054 32153: gap of 100 bp  
 32154 32179: contig of 3026 bp in length  
 32180 32279: gap of 100 bp  
 32280 32448: contig of 2969 bp in length  
 32449 33548: gap of 100 bp  
 33549 33847: contig of 3499 bp in length  
 33848 33947: gap of 100 bp  
 33949 34073: contig of 4126 bp in length  
 34074 43173: gap of 100 bp  
 43174 47538: contig of 4365 bp in length  
 47539 47638: gap of 100 bp  
 47639 53052: contig of 5414 bp in length  
 53053 53152: gap of 100 bp  
 53153 58427: contig of 5275 bp in length  
 58428 58527: gap of 100 bp  
 58528 67461: contig of 8934 bp in length  
 67462 79052: contig of 11491 bp in length  
 79053 92165: contig of 13013 bp in length  
 92166 92265: gap of 100 bp  
 92266 110031: contig of 17766 bp in length  
 110032 110131: gap of 100 bp  
 110132 129806: contig of 19675 bp in length  
 129807 129906: gap of 100 bp  
 129907 150129: contig of 20223 bp in length.

FEATURES

	Location/Qualifiers	Source
	1..150129	/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/chromosome="1"
		/map="1"
		/clone_lib="RPC1-11 Human Male BAC"
misc_feature	1..1572	/note="assembly_fragment"
misc_feature	1673..2741	/note="assembly_fragment"
misc_feature	2842..4870	/note="assembly_fragment"
misc_feature	4971..6056	/note="assembly_fragment"
misc_feature	6157..7419	/note="assembly_fragment"
misc_feature	7520..9305	/note="assembly_fragment"
misc_feature	9406..11199	/note="assembly_fragment"
misc_feature	11300..13054	/note="assembly_fragment"
misc_feature	13155..15817	/note="assembly_fragment"
misc_feature	15918..18265	/note="assembly_fragment"
misc_feature	18366..20797	/note="assembly_fragment"
misc_feature	20898..24397	/note="assembly_fragment"
misc_feature	24498..27417	/note="assembly_fragment"
misc_feature	27518..29053	/note="assembly_fragment"
misc_feature	29154..32179	/note="assembly_fragment"
misc_feature	32280..35248	/note="assembly_fragment"
misc_feature	35349..38847	/note="assembly_fragment"
misc_feature	38948..43073	/note="assembly_fragment"
misc_feature	43174..47538	/note="assembly_fragment"

```

/misc_feature /note="assembly_fragment"
47639..53052
/misc_feature /note="assembly_fragment"
53153..58427
/misc_feature /note="assembly_fragment"
58528..67461
/misc_feature /note="assembly_fragment"
67562..79052
/misc_feature /note="assembly_fragment"
79153..92165
/misc_feature /note="assembly_fragment"
92266..110031
/misc_feature /note="assembly_fragment"
clone_end:SP6
vector_side:left"
110132..129806
/misc_feature /note="assembly_fragment"
129907..150129
/misc_feature /note="assembly_fragment"

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 9,91e-210 Length: 150129
Score: 3333.00 Matches: 632
Percent Similarity: 96.95% Conservative: 3
Best Local Similarity: 96.49% Mismatches: 8
Query Match: 47.68% Indels: 12
DB: 2 Gaps: 2

```

US-09-744-167-2 (1-1323) x AC026909 (1-150129)

```

QY 1 MetTpIleAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrSerTirp 20
DB 70313 ATGTGGATTGATGAATGCTGTTCGACAGACCCAGTTAATTAAGAAACCTATAGTTGG 70372
QY 21 AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu 40
DB 70373 GATGATCAATGAGTGTCTGTTGAAGTGGAGAGAAATGTGGAAACCTGGCTGTCTG 70432
QY 41 ProAspGluLysAsnValLeuValAlaValAlaValMethIshenCysAspLysArgThrLeu 60
DB 70433 CCAGATGAGAGAAATGTTCTGTGTAGCCGCTCATGATCAATCTGTGATTAAGGACATTA 70492
QY 61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80
DB 70493 CAAACCGATTACAGGATTGTAATTAATTAATAGTCAATCCCTTATGGATGCTTTTAGC 70552
QY 81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100
DB 70553 TGTTCACTGGATAATGAACAGACAAACTGATCAATTTAGTTTATAGTATAAATGAGTCC 70612
QY 101 ThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu 120
DB 70613 ACTGAAAGATATGATTCAGAGAAACAAATGATCCATTCATAGACCCAAACAGAG 70672
QY 121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140
DB 70673 GGGAGATCTGTTAACCATCTGTGTCTTACTTCATCTCATCTAGTCTAGCCAGTGTCTGTCC 70732
QY 141 ProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160
DB 70733 CCTTCACAAATTAAGGATGACGGAAGATAGTAGAGACCCCTCCCATGCTCGGATTACA 70792
QY 161 SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys 180
DB 70793 AGTTTAACCGTTGATTGATTAATCTCATCCAGGAACAGATGGATGTCCTGCTGTATAA 70852
QY 181 LysGlnGluAsnTyrIleProAspGluAspLeuThrClyLysIleSerSerProArgThr 200
DB 70853 AAGCAGAGAACTATATACAGATGAGGACCTCAGTGGCAAAATCAGCTCTCCCTAGGACA 70912
QY 201 AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu 220

```

```

DB 70913 GATCTAGGGAGTCCAAATTCCTTTTCCACATGAGTGGGGATTTTGATGAAAAAGAG 70972
QY 221 ProAlaGluGluSerThrThrGluGluSerLeuArgSerGlyLeuProLeuLeuLys 240
DB 70973 CCAGCAGAGAGAGACACCACTGAGAAATCCCTCCGGTCTCGGTTTACCTTTGCTTCAAA 71032
QY 241 ProAspMetProAsnGlySerGlyArgAsnAsnAspCysGluArgCysSerAspCysLeu 260
DB 71033 CCAGACATGCCTAATGGGTCTGCAAGGAATAATGACTGTGAACGGTGTTCAGATTGCCTT 71092
QY 261 ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly 280
DB 71093 GTGCCCTAATGAAGTTAGGGCTGATGAATGAAGGTATGAACATGAGAAACTCTTGCC 71152
QY 281 ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTirp 300
DB 71153 ACTACAGAATTCCTTAATATGACAGAGCATTTCTCTGAATCTCAGACATGACTAATTGG 71212
QY 301 LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluGluLysGluLysPhe 320
DB 71213 AAGTTGACTAAACTAAATGAGATGAATGATAGCCAAAGTAAACGAAAGAAAGAAATTT 71272
QY 321 LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlyGlnCysValGlyLeu 340
DB 71273 CTACAGATTAGTCAGCCTGAGGACACTAATGTGTAGTGGAGGACAGTGTGTGGATTG 71332
QY 341 AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe 360
DB 71333 GCAGATGCAGGTCTAGATTAAAGGAACTTGCCATTAGTAAAGTGAAGATGTGATTTC 71392
QY 361 SerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCysAspSerTyrGly 380
DB 71393 TCACCTGTTATAGACACACCAGCAGCAATATATCTATCTAATGTTGTGATTCCTATGGA 71452
QY 381 MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal 400
DB 71453 ATGCAAGACCCAGGTGTTCTTTTCTCCAAAGACTTTTACCTCCAAAGAAAGATTCACTA 71512
QY 401 ThrGluGluLysGluIleGluGluSerLysSerGluCysTyrSerAsnIleTyrGluGln 420
DB 71513 ACAGAGAAAGAAAGAAATAGAGAAAGCAAGTCAAGATCTACTCAATATATTATGAACAG 71572
QY 421 ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuLeuAsnSerThrGlyAspLeuMet 440
DB 71573 AGAGGAATATGAGCCACAGAGAGGAGTGGACTACTTTTAAACAGCACTGGTGACCTAATG 71632
QY 441 LysLysAsnTyrLeuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer 460
DB 71633 AAGAAAAATTAATTTATACAAATTTCTGTAGTCAAGTTCATCAGTGTCTGGCAATCTTCC 71692
QY 461 ProLysValValAlaSerLeuProSerIleSerValPropheGlyGlyAlaArgProLys 480
DB 71693 CCCAAGGTAGTAGCAGCCTGCCATCTATCAGTGTCTCTTTTGGTGTGCAAGACCCAG 71752
QY 481 GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp 500
DB 71753 CAACCTCTTAATCTTAAACTTTCAAAATTCAAAGCCATTTATCAGACCATTTTACAAATGAC 71812
QY 501 PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys 520
DB 71813 TTTCTCGCAACAGTGGAAATTAATTAATAAATAAATGATATTTCTTGGGAAGCAAAA 71872
QY 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540
DB 71873 TTAGGGGAAAACTCAGCAACCAATGTATGCACTCCATCTTTTGGGAAACATCTCTTAATGTC 71932
QY 541 AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu 560
DB 71933 GATACAAATGGGGAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTT 71992
QY 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580
DB 71993 GCATTAGTCCAGATAGCCCAAGATAATGATCTCAGAGCTGGTCAGTGTGGAAATTTCTGCC 72052

```

```

Qy 581 ArgLysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla 600
Db 72053 AGAAGGCATTCACCACTCTGGGTGGAGTGGTCCAGTATGGTACCGGATTCAGGCT 72112
Qy 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisHisCysArg 620
Db 72113 CCANATTCATGAATGTCGAAGCAGGTTTACATTCACCAAGAGGAGGATCTACTGAGA 72172
Qy 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTyrMet 640
Db 72173 GCATGTGGGAAGGTAACTGC-----ATGTATACG 72202
Qy 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeu 655
Db 72203 CTCAGAAAT-----CGGGATGCACATTTTGTAATGCTGAATTA 72241

RESULT 9
AL591398 89973 bp DNA linear PRI 06-JUN-2001
LOCUS Human DNA sequence from clone RP11-81C12 on chromosome X, complete
DEFINITION
ACCESSION AL591398
VERSION AL591398.2 GI:143330165
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Heath,P.
1 (bases 1 to 89973)
Direct Submission
Submitted (05-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced gi:14133279.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
Chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP11-81C12 is from the library RP11-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-81C12 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP1-389B13 is at 89874 in this sequence.
The true right end of clone RP1-137H15 is at 100 in this sequence.
The true right end of clone RP1-7706 is at 44281 in this sequence.
Location/Qualifiers
1..89973
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
FEATURES
source
/chromosome="X"
/clone="RP11-81C12"
/clone_lib="RPC1-11.1"
1..7991
/note="LiMA2 repeat: matches 5524..6308 of consensus"
914..1079
/note="LiMD repeat: matches 994..1160 of consensus"
1130..1329
/note="LiMEC repeat: matches 2218..2411 of consensus"
1330..1377
/note="LiMTA1 repeat: matches 318..365 of consensus"
1378..1680
/note="AluO repeat: matches 1..308 of consensus"
1681..1848
/note="LiMTA1 repeat: matches 155..318 of consensus"
1849..2152
/note="AluX repeat: matches 1..299 of consensus"
2153..2305
/note="LiMTA1 repeat: matches 1..155 of consensus"
2306..2602
/note="LiMEC repeat: matches 1901..2218 of consensus"
2642..3475
/note="LiMEC repeat: matches 1245..2090 of consensus"
3476..3860
/note="THE1C repeat: matches 3..371 of consensus"
3861..3929
/note="LiMEC repeat: matches 2090..2159 of consensus"
3930..4240
/note="AluY repeat: matches 1..306 of consensus"
4241..4885
/note="LiMEC repeat: matches 2159..2430 of consensus"
4886..5198
/note="AluJb repeat: matches 1..309 of consensus"
5199..5491
/note="LiMEC repeat: matches 2430..2726 of consensus"
5491..5917
/note="LiM4 repeat: matches 3146..3578 of consensus"
5918..6235
/note="AluJb repeat: matches 1..312 of consensus"
6236..6976
/note="LiM4 repeat: matches 3578..4297 of consensus"
7047..7160
/note="Charliel repeat: matches 2022..2140 of consensus"
7161..7262
/note="LiM3 repeat: matches 7440..7542 of consensus"
7263..7581
/note="AluX repeat: matches 1..302 of consensus"
7582..7776
/note="LiM3 repeat: matches 7542..7734 of consensus"
7777..7791
/note="Charliel repeat: matches 2140..2153 of consensus"
8810..8960
/note="Charliel repeat: matches 2610..2761 of consensus"
9033..9458
/note="Charliel repeat: matches 2190..2627 of consensus"
9506..9826
/note="AluSq repeat: matches 1..308 of consensus"
9938..10017
/note="AluS repeat: matches 1..80 of consensus"
10019..10166
/note="LiM3 repeat: matches 6611..6747 of consensus"
10205..10502
/note="AluX repeat: matches 1..311 of consensus"
10538..10681
/note="LiM3 repeat: matches 6468..6609 of consensus"
11563..11712
/note="FAM repeat: matches 13..165 of consensus"
11906..12226
/note="AluX repeat: matches 1..312 of consensus"
12345..12652
/note="AluSp repeat: matches 1..310 of consensus"
12992..13267
/note="AluJb repeat: matches 1..285 of consensus"

```



Db 52883 AAAAATTCCCAACACATATAAATGAGATGATGAAAGCCACGAAACCAAGTCCATGACGA 52942

Qy 1055 sValleualaglyGlyAlaCysPheAsnGluLysAlaAspSerHisLeuValCysValG1 1075

Db 52943 TATCTTGGCAGAGGTGCTCTCTCAATGAAAGGACGACTCTCATCTTGTGTGTA--- 52999

Qy 1075 nAsnAspGlyAsnTyrGlnThrGlnAlaIleSerIleHisAsnGlnProArgLysVa 1095

Db 53000 -----AGCTATCATGAGCCAGGCTATCATGATTATCACAATCAGCCAGAAAAGT 53046

Qy 1095 lThrGlyAlaSerPhePheValPheSerGlyAlaLeuLysSerSerSerGlyTyrLeuAl 1115

Db 53047 GACTGGTGCCAGTCTCTTTGTTCAGTGGCAGCTCTGAAATACATCTTCTGGAACCTTGC 53106

Qy 1115 aLysSerSerIleValGluAspGlyValMetValGlnIleThrAlaGluAsnMetAspSe 1135

Db 53107 CAAGTCAGATATTGGGAAGATGGGTATGTGTCACAGATCACTGCAGAAACATGGATTG 53166

Qy 1135 rLeuArgGlnAlaLeuArgGluMetLysAspPheThrIleThrCysGlyLysAlaAspAl 1155

Db 53167 CTTGAGGAGGCACTGTGAGAGATGAAGGACTTCACCATCACCCGTGGGAGGCGCACGC 53226

Qy 1155 aGluGluProGlnGluHisIleHisIleGlnTyrValAspAspLysAsnValSerLy 1175

Db 53227 AGAGATCCCCGAGGACATCCACACCAATGGTGGATGATGACAGAACGTTAGCAA 53286

Qy 1175 sGlyValValSerProLysAspGlyLysSerMetGluThrIleThrAsnValLysIlePh 1195

Db 53287 GGGCGTGTAAAGTCTATAGATGGGAAGTCCATGGAGACTATAACAATGTGAAGATATT 53346

Qy 1195 eHisGlySerGluTyrLysAlaAsnGlyLysValIleArgTyrThrGluValPhePheLe 1215

Db 53347 CTATGGATCAGAAATGAAGCAATGAAGAAAGTATCATGATGGACAGAGGTGTTTTCCC 53406

Qy 1215 uGluAsnAspGlnHisAsnCysLeuSerAspProAlaAspHisSerArgLeuThrG1 1235

Db 53407 AGAAATCCTCACCAGCACAAATGGCTCAGTGATCTCGAGATCACAGTAGATTGACTGA 53466

Qy 1235 uHisValAlaLysAlaPheCysLeuAlaLeuCysThrGlnLeuLysLeuLysGlyAs 1255

Db 53467 GCATGTTTACCAGGCTTTTGTCTGTCTCTGCTCCTCACCTGAAGCTCTGGAAGGA 53526

Qy 1255 pGlyMetThrLysLeuGlyLeuArgValThrLeuAspSerAspGlnValGlyTyrGlnAl 1275

Db 53527 TGGATGACCAAACTGGGACTACGTGTGACACTTGTATTCAGATCAGTCTGGCTATCAAGC 53586

Qy 1275 aGlySerAsnGlyGlnHisLeuProSerGlnTyrMetAsnAspPheAspSerAspLeuVa 1295

Db 53587 AGGGAGAAATGGCCAGGCCCCCTCCCTCGCACTGCATGAATGATCTGCACAGCGCCTGGT 53646

Qy 1295 lLysMetIleHisGlyGlyAlaCysGlnLeuSerGluGlyProValValMetGluLeu11 1315

Db 53647 GCCGTGATCTCATGGAGGGGCTGTGCTCAGTCTGAGGGCTCTGTCTGATGGAACATCAT 53706

Qy 1315 e 1315

Db 53707 T 53707

RESULT 10

AL671909/c

LOCUS Mouse DNA sequence from clone RP23-446F20 on chromosome 4, complete sequence.

DEFINITION

ACCESSION AL671909

VERSION AL671909.8 GI:22797964

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS Brown, J.

TITLE Direct Submission

## JOURNAL

## COMMENT

Submitted (06-SEP-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Sep 12, 2002 this sequence version replaced gi:2121224. clones. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-446F20 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACes.6

## -----

Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

## FEATURES

## source

1. 85802

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="4"

/clone="RP23-446F20"

/clone\_lib="RPC1-23"

## ORIGIN

## Alignment Scores:

Pred. No.:	1,06e-152	Length:	85802
Score:	2471.50	Matches:	482
Percent Similarity:	84.37%	Conservative:	47
Best Local Similarity:	76.87%	Mismatches:	71
Query Match:	35.36%	Indels:	27
DB:	10	Gaps:	6

US-09-744-167-2 (1-1323) x AL671909 (1-85802)

Qy	1	MetTrrPileAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrSerTrp	20
Db	60423	ATGTGGATTGATGAAATCTGTGACAGAGCAACCACTAATTAAGAGAACTATATACAG	60364
Qy	21	AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu	40
Db	60363	GATGATCAATTTAGTGCCTTGAAGTGGAGAGGAGAACTGGAAGCTCACTTGTCTG	60304
Qy	41	ProAspGluLysAsnValLeuValAlaValMetHisAsnCysAspLysArgThrLeu	60
Db	60303	CCAGACGAGAAGAAATGTTCTGTGTGTAGCTGTGATGATGATGATGATGATGATG	60244
Qy	61	GlnAsnAspLeuGluAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer	80
Db	60243	CAAGCGATTTTCAGGATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT	60184
Qy	81	CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer	100
Db	60183	TGTTCACTGGATAATCAGACTAGACAACTGATCAATTTAGTTAGTATGATGATGATG	60124
Qy	101	ThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu	120







Db	181881	TCCTTCAAAATTAAGGACAAATGAAATGATAGTACAGCCATCTACATCTACAGCTACA	181822	Db	180882	CTTCTCGCAACAATGGAATAATAGTAAAAAATAATGATATCTTGGGAAGCAAAA	180823
Qy	161	SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys	180	Qy	521	LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal	540
Db	181821	AGCTTAGCAATTAAT-----ACATCCAGGGAATGATGAGGCGCTGCACTAAA	181771	Db	180822	CTAGGGGAAAATCTCAGCAGTCAGTGCAGTCAACTTTGGGAAAACATCTCTGTGACT	180763
Qy	181	LysGlnGluAsnTyrlleProAspGluAspLeuThrGlyLysIleSerSerProArgThr	200	Qy	541	AspThrAsnGlyGluHisLeuGlySerTyrGluAlaGluIleSerThrArgProCysLeu	560
Db	181770	AAACAAGAGGATTATATGCTGCTGATGAGACCTTTCTGGCGCAACACAGCTCTCTAGGACA	181711	Db	180762	GATACAAATGGGGAACATTTAGAAAGTTTCGAGGCTGCATCTCCAGTAGCACTGCTT	180703
Qy	201	AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu	220	Qy	561	AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla	580
Db	181710	GATCTAGGGAGCTCAAACTCTCTTCCCACTCAAGGAGGAGCTCTTAAACAAAACACAG	181651	Db	180702	GCATTATCTCCAGATAGTCCAGATAATGATCTCAGAGCTGCTCAATTGGGAATTTCTGCA	180643
Qy	221	ProAlaGluGluSerThrThrGluGluSerLeuArgSerGlyLeuProLeuLeuLys	240	Qy	581	ArgLysProPheThrThrLeuGlyGluValAlaProValThrValProAspSerGlnAla	600
Db	181650	CCAGCAGAGGAGAGACAGCTGAGACCCCTCCATCTGGTTATCTTTGAATCTCAAG	181591	Db	180642	AGAAAGCCATTTTACCACCTCGGTGAGTGGCAGCAGTATGGTACCAAGATCCAGGCT	180583
Qy	241	ProAspMetProAsnGlySerGlyArgAsnAsnAspCysGluArgCysSerAspCysLeu	260	Qy	601	ProAsnCysMetLysCysGluAlaArgPheThrLysArgArgHisCysArg	620
Db	181590	CCAGATACCTCTGCTTTATGTGGAGGACAAT---TGCGAACCATCTCAGACTGCTT	181534	Db	180582	CCGAATCTGCATGACGTGAGCTGAGCTCGTTTACATTCACCAAGAGGAGGCATCATTCGAGA	180523
Qy	261	ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly	280	Qy	621	AlaCysGlyLysValPheCys	627
Db	181533	GTGCTAGTAGACTGAGACTGATGGAGATGGAGT-----	181498	Db	180522	GCATGTGGGAGGTAAATTGC	180502
Qy	281	ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTrp	300	RESULT 12			
Db	181497	-----AAGGACTCTCAGCAGACTCACTGG	181474	AC118106/c			
Qy	301	LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluGluLysGluLysPhe	320	LOCUS			
Db	181473	GACTCAGCAAGCTAAATGAGTGGATGATCCCAAGACAGTGAAGAAATCAAGTGGT	181414	DEFINITION			
Qy	321	LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlnCysValGlyLeu	340	AC118106			
Db	181413	TTACAGATCCCAACTGAGACCT-----CAGAGTGCAGGAGAGTGTGTAGAATG	181360	VERSION			
Qy	341	AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe	360	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.			
Db	181359	CGCAGATTCCAGATCTAGATTTCAAAGGAACTTGTGTAAGTAAGTGAAGGATATGATTC	181300	KEYWORDS			
Qy	361	SerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCysAspSerTyrGly	380	SOURCE			
Db	181299	TCCACTGTAGTGATGCACAGCAGCAAAATCTCTGCTTAATAGTCTGATCTCTATGGA	181240	ORGANISM			
Qy	381	MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal	400	REFERENCE			
Db	181239	ATGCAGAGCCCAATGTATCTTTGTTCCAAAGACTTTACCTCCAAAGAGATTCAGTA	181180	AUTHORS			
Qy	401	ThrGluGluLysGluIleGluSerLysSerGluCysTyrSerAsnIleTyrGluGln	420				
Db	181179	ACAGAAGAAAAGAAATAGAGAGAGCAAAATCAGAAATGCTACTCAATCTTTATGAACAG	181120				
Qy	421	ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuLeuAsnSerThrGlyAspLeuMet	440				
Db	181119	AGAGGAAATGAGAACACAGAGGGAATGCACTGCTTTTAAACAGCACTGGAACTATAATG	181060				
Qy	441	LysLysAsnTyrLeuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer	460				
Db	181059	AAGAAAAATTTATTCATAAATTCGTAGTCAAAATCCCATCTGCTACTCGGGCAATCTCC	181000				
Qy	461	ProLysValAlaSerLeuProSerIleSerValProPheGlyGlyAlaArgProLys	480				
Db	180999	CCCAAG---ATAGCAAACTGCAGCTATTAGTGTCTCTTTTGTGGTGCAGACCCCAAG	180943				
Qy	481	GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp	500				
Db	180942	CAACCTCTTAATCTTAAACTTCAGATTCCAAGGCAATTCATCAGACCAATTCACAAAATGAC	180883				
Qy	501	PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys	520				

AC118106 201330 bp DNA linear HTG 13-NOV-2002  
Rattus norvegicus clone CH230-324B23, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 2 unordered pieces.

AC118106  
AC118106.5 GI:24941760  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

1 (bases 1 to 201330)  
Muzny, D. Marie, Metsker, M. Lee, Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebreragis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hui, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenshew, L., Louisedge, H., Lozard, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, N., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, N., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plopper, F., Poindexter, A., Popovic, D., Reeves, K., Regier, M.A., Reigh, R., Puazo, M., Quiroz, J., Rachlin, E., Reuter, M., Richards, S., Riggs, F., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Rose, R., Ruiz, S.J., Rivers, C., Rodey, I., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savory, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

# TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

## REFERENCE AUTHORS TITLE JOURNAL

### COMMENT

Submitted (13-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 201330)

Rat Genome Sequencing Consortium.

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 13, 2002 this sequence version replaced gi:23194893.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GVED  
Center clone name: CH230-324B23

----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 188149 bases at least Q40  
Consensus quality: 190201 bases at least Q30  
Consensus quality: 191458 bases at least Q20

Estimated insert size: 194406; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

-----  
\* 1 198699: contig of 198699 bp in length  
\* 198700 198799: gap of unknown length  
\* 198800 201330: contig of 2531 bp in length.  
\* Location/Qualifiers

### FEATURES

source 1..201330  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-324B23"  
1..1448  
/note="wgs end extension  
clone\_end:Sp6"  
1920..4036  
/note="wgs end extension  
clone\_end:Sp6"  
complement(6631..7552)  
/note="clone boundary  
clone\_end:Sp6  
site\_end:Sp6  
end\_sequence:BZ238666"

ORIGIN

Alignment Scores:  
Pred. No.: 1.69e-124 Length: 201330  
Score: 2056.00 Matches: 416  
Percent Similarity: 74.80% Conservative: 53  
Best Local Similarity: 66.35% Mismatches: 131  
Query Match: 29.41% Indels: 28  
DB: 2 Gaps: 5

US-09-744-167-2 (1-1323) x AC118106 (1-201330)

QY 1 MetTPIIAspGluAsnAlaValAlaGluAspGlnLeuLeuIlelyAsnTyrSerTrp 20  
Db 34915 ATGTGGATTGATGAAATGCTGTGGCAGAGCAACCACTAATTAAGAGAACTACATCAG 34856

QY 21 AspAspGlnCysSerAlaValGluValGluGlyGluLysCysGlyAsnLeuAlaCysLeu 40  
Db 34855 GATGATCAATTTAGTGGCTTGAAGTGGAGAGAGAAACGTGGGAGCCTCACTTGTCTG 34796

QY 41 ProAspGluLysAsnValLeuValValAlaValMetHisAsnCysAspLysArgThrLeu 60  
Db 34795 CCAGACGAGAGAAGATGTTCTGGTTGTAGCTGCATGCATAACTGATAAAAGGACACTG 34736

QY 61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80  
Db 34735 CAAGTGATTGACGAGATTGTAATTAATATATATATATATATATATATATATATATATAT 34676

QY 81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100  
Db 34675 TGTTCACTGGATAATGAAACTAGACAACTAGTCAATTTAGTTTATGATGAATGGGTCC 34616

QY 101 ThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu 120  
Db 34615 GCAGAAAAGGTATTAACCTCAGAGAAACAAATGGATGCTTTGAAATAAACCAAAATCAGAG 34556

QY 121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140  
Db 34555 AGGGATTCTCTTAATCACTCTCGCTTCAGCTAACAGTGCAACAGTATCTCTTCC 34496

QY 141 ProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160  
Db 34495 TCTTCACAAATTAAGACAACTGAAATGTAGTAGAGACCCATCATCTACAGCTACA 34436

QY 161 SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys 180  
Db 34435 AGCTTAGCAATTAAT-----ACATCCCAAGGAATGGATGGAGGCGCTGCCATTAAA 34385

QY 181 LysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysIleSerSerProArgThr 200  
Db 34384 AAACAAGAGGATTATATGCTTCATCAGGACCTTTCTGGCGCAACAGCTCTTCTAGGACA 34325

QY 201 AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysGlu 220  
Db 34324 GATCTAGGAGCTCAAACTCTCTTTCCCACTCAAGGAGGAGCTCTTAACAAAAACAGAG 34265

QY 221 ProAlaGluGluSerThrThrGluGluSerLeuArgSerGlyLeuProLeuLeuLys 240



## FEATURES

Location/Qualifiers  
 1..5480  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="MGC:28649 IMAGE:4235024"  
 /tissue type="Kidney, normal, 5 month old male mouse."  
 /clone\_lib="NCI CGAP\_Kid14"  
 /lab\_host="DH10B"  
 /note="vector: pCMV-SPORT6"  
 1..5480  
 /gene="B130024H06Rik"  
 /note="synonym: B130031I15"  
 /db\_xref="LocusID:218441"  
 /db\_xref="MGI:2145181"  
 150..4736  
 /codon\_start=1  
 /product="RIKEN cDNA B130024H06"  
 /protein\_id="AAH42669.1"  
 /db\_xref="GI:27503686"  
 /db\_xref="LocusID:218441"  
 /translation="MDSYFYAAVSGLDKLLDDPEONPKDYLDPAYAFNCQSVSES  
 ASPQALLSDQRCISACSEACEDANETFLGKIHEGLTRSRNEKNVAGDDLLSS  
 VDASTDEIQSCMRCKSPVCDLSDMGLNATNSEEIDIKQLLPDDPKSSADTLIT  
 LDSSVSEALTVSSVDCGNAVREEQNNINAGIKREDISIKELGVYDVALFDCSKYN  
 RTENLKDYLISNELETYDFDMPVSLMEQSEMGNKNDKQYKRLPCELLKDDCLAE  
 KVAVANVTECLERGGSNTIAMPKLPENEGISPSDPAASKDENFKLPDPLOENRTS  
 VFMKQYKEDSRNLDKNDNDIVHVSDDVPVPSLCSLSGSLGSLIHNNHSDILP  
 PSESGONNADVAETHEIKQSDVLDETDLSKKECHSIPLQVNEKKCKGKVEVEM  
 VISGESLESDEASAAAGSPVALSAASPEAPGPCGELTFPSSMDGQELDYFNID  
 ESMKSGILISDAELDAFLKEQCSNNTMSAGENVDSQMQNQTIMKGLHNDAGDI  
 YFNAAGAAGENGVCNLTSDKENTENGLSIGERKGAIFTERELSACTVSTCSADHP  
 PSITQAVGARPKQLISLPGRSSKELNKPVDVDPSEPCATANATVSTCSADHP  
 PDSOVNSNVYDIENFEDGSPFTANKDSLLENKREKSLVIGOKPTWVPDSEAPN  
 CMACOVATKRRHHCACQKVFQVGNCKKLOVLEKEARVCVCTINKAQAF  
 ERMSPGSGCLKNSNECATDQPLQETQTSPTSPPTLPIALKQPNVEGPKSKQK  
 RWFDAGILNGEVADTILKSSGKRCDDFSPVLPDVPMTINKVORTHSPTVEKNP  
 GLGDIIEISIQCTHAPVRLPGNTGTEGLPMGPFTLEDDVFDVDSSEPTTWW  
 PANSGLPVASISDRLLCVSAKVCNNISLLPDDIGLPLLATSGEDSGSPVYQRRP  
 SHEQIILLGEGFPFATFVNANLVNKLVLVSSEKYYWFTNGHLGQAEITVL  
 LQCPNEDTPKDFITITIKDALGKVIENLDNLTFTFESFLNSKHGGLFIPT  
 FONDGLPVRSPLGKILQLEIPKAKVPMRLMLRLGAEYKAYPAPLTSVGRKPK  
 LFGEIGHTMNLVDLRNYQVTLNIDQLLIHMEGKSCIKIPRKXYSVMKVIHSN  
 EHVISGASFEADSHLVQSDGVYQANSATQPKRVGTGASVVFVNGAUKTSNG  
 FLAKSIVELQVQIPTETMEGLRLALREKQFRIQCQKVDADLVREYDVCWDS  
 ERKNKGVSSVDGMSVGFPSKIKLETFETBEKTVKTEVFYFLKDDQISLSSY  
 QFAKEIAVACSAALCPHLRLKSNRNKIGLRVSDIDMVEFQAGCGQLLPQHYLND  
 LDSALIPVINGTSNSSLPLEIELAPFLENLSE"  
 2337..2531  
 /note="FYVE; Region: FYVE zinc finger. The FYVE zinc  
 finger is named after four proteins that it has been found  
 in: Fab1, YOTB/ZK632.12, Vac1, and EEA1. The FYVE finger  
 has been shown to bind two Zn++ ions. The FYVE finger has  
 eight potential zinc coordinating cysteine positions. Many  
 members of this family also include two histidines in a  
 motif R-HHC+XCG, where + represents a charged residue and  
 X any residue. We have included members which do not  
 conserve these histidine residues but are clearly related"  
 /db\_xref="CDD:pfam01363"

## gene

1..5480  
 /gene="B130024H06Rik"  
 /note="synonym: B130031I15"  
 /db\_xref="LocusID:218441"  
 /db\_xref="MGI:2145181"

## CDS

150..4736  
 /codon\_start=1  
 /product="RIKEN cDNA B130024H06"  
 /protein\_id="AAH42669.1"  
 /db\_xref="GI:27503686"  
 /db\_xref="LocusID:218441"

## misc\_feature

2337..2531  
 /note="FYVE; Region: FYVE zinc finger. The FYVE zinc  
 finger is named after four proteins that it has been found  
 in: Fab1, YOTB/ZK632.12, Vac1, and EEA1. The FYVE finger  
 has been shown to bind two Zn++ ions. The FYVE finger has  
 eight potential zinc coordinating cysteine positions. Many  
 members of this family also include two histidines in a  
 motif R-HHC+XCG, where + represents a charged residue and  
 X any residue. We have included members which do not  
 conserve these histidine residues but are clearly related"  
 /db\_xref="CDD:pfam01363"

## ORIGIN

## Alignment Scores:

Pred. No.: 3,86e-126 Length: 5480  
 Score: 2047.00 Matches: 523  
 Percent Similarity: 49.56% Conservative: 208  
 Best Local Similarity: 35.46% Mismatches: 481  
 Query Match: 29.28% Indels: 264  
 DB: 10 Gaps: 40

US-09-744-167-2 (1-1323) x BC042669 (1-5480)

Qy 3 IleAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTrpSerTrpAspAsp 22

Db 630 TTGGACTGCTCTTCAGTTTCAGAG----- 653  
 Qy 23 GlnCysSerAlaValGluValGlyGluLysLysCysGlyAsnLeuAlaCysLeuProAsp 42  
 Db 654 -----GCTCTCACTGCTCTCTCAGTAGACTGTGGTAGTAATGCTGTACAGAGAG 704  
 Qy 43 GluLysAsnValLeuValAlaValMetHisAsnCysAspLysArgThrLeuGlnAsn 62  
 Db 705 CAGATAACATC-----AACGCTGGTATATAAAACAGACATATT 743  
 Qy 63 AspLeuGlnAspCysAsnAsnTrpAsnSerGlnSerLeuMetAspAlaPheSerCysSer 82  
 Db 744 AGTATCAAGAATGGCTGTAAAGTAGACATGGCACTTTTTCAT-----TCTTCTAAA 797  
 Qy 83 LeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSerThrGlu 102  
 Db 798 TATAATAGAACAGAAATTTAAAGAGATAAAATTTCTAATGAGTTAGTAAACAGTTGAT 857  
 Qy 103 LysAspMetAsnSer-----GluLysGlnMetAspProLeuAsnArgProLysThrGlu 120  
 Db 858 TTTGACATCCCATCTGTTTGTGATGGAACAAAGTTCTGAAATGCTCTAATACCAAGATAAC 917  
 Qy 121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140  
 Db 918 CCACAATACAGAGGCTGCCATGT----- 941  
 Qy 141 ProSerGlnLeuLysAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160  
 Db 942 ---GAATTCTTAAAGATGATGCTGTTGGCAGAGAGAAAGTAGCTGCGCAGTCAAT 998  
 Qy 161 -----SerLeuThrValAsp-----SerVal 167  
 Db 999 AATACAGAGTGTTTAGAGAGAGAGAGAGAGTAAACAAATTCGTCTGCTTGCACAGCTT 1058  
 Qy 168 IleSerSerGlnGlyThrAspGlyCysProAlaValLysLysGlnLysThr----- 186  
 Db 1059 CAGAAAATGAAGGATATCCCAAGTAGCCAGCTTCAAAAGATGAAAATTTCAAGTTA 1118  
 Qy 187 ProAspGluAspLeu-----ThrGlyLysIle 195  
 Db 1119 CQTGACTTTCCCTCCAGGAAATAGAACTTCTGTGTTTCATGAAACAGACTGTAAAGAA 1178  
 Qy 196 SerSerProArgThrAspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIle 215  
 Db 1179 GACTCAAGAAATTTAGACCTTAAAGATAATAATGATATAGTCCATGTTTCA----- 1229  
 Qy 216 LeuMetLysLysGluProAlaGluSerThrThrGluLysLeuSerLeuArgSerGlyLeu 235  
 Db 1230 -----GGTGATCATGTGCCCTTCACTGCTCTTCTTCTTCTGCTGGCTTTG 1280  
 Qy 236 ProLeuLeuLysProAspMetProAsnGlySerGlyArgAsnAspCysGluArg 255  
 Db 1281 -----TGTGGATCGTTAAATTCACATAAT-----GAA 1307  
 Qy 256 CysSerAspCysLeuValProAsnGluValArgAlaAspGluAsnGluGlyTyrgluHis 275  
 Db 1308 CACAGTATATTTTCTCTCTATGAGTCTGAAGCCAGACAAATGATGAGTACTATA 1367  
 Qy 276 GluGluThrLeuGlyThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGln 295  
 Db 1368 CATGAAGACATACAGAAGAGTATGTTTTAGATGGGAGACAGATCTTTCAAAAGAGGAA 1427  
 Qy 296 AspMetThrAsnTrpLysLeuLysLeuAsnGluMetAsnAspSerGlnValAsnGlu 315  
 Db 1428 AGCTGTAGAGCATATTTCTTCAGCCAGTAAATGAAAAGAGGGAGGAGGAAAGTTGAG 1487  
 Qy 316 GluLysGluLysPheLeu-----GlnIleSerGlnProGluAspThrAsnGly--- 331  
 Db 1488 GTTGAGGATGATGTAATCAGTGTGAGTCTTTGGAGTCCCTCTGAGGACGCCAGCTCTGCT 1547  
 Qy 332 -----AspSerGlyGln 336

1548	Db	GCAGCAGCTGGGTCTCCGGTGGCACTTTCTGTGTCGAAGTGTCCAGAGGCGTCTCTGTGTC	1607
337	Qy	CysValGlyLeuAlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGlu	356
1608	Db	TGTGAAGGTCCTCACTTTTCCAGCAGTGATGGATGGG	1649
357	Qy	GluCysAspPheSerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCys	376
1650	Db	GAGTTAGATTACTTAAATATGTGAAGAGATGAGAAGTGGCATACATAATTAGCGA-TGC	1708
377	Qy	AspSerTyrGlyMetGlnAspProGlyValSerPheVal-ProLysIleThrLeuProSerLy	396
1709	Db	TGAACTCGATGCTTTCTGAAGGAACAGTGTCTTCGAACCTCTAACACAATGCTGCTGG	1768
396	Qy	sGlu-----AspSerValThrGluGluLysGluIle-----G1	407
1769	Db	AGAAAAATGAATGATTTCCCGATTGCGAGTGAATCAGATCACTATGAAGAGCTACATGA	1828
407	Qy	uGluSerLysSerGluCysTyrSerAsnIle-----	417
1829	Db	TGAATATGCTGGGATATATCTTCAATGCTGAAGCAGGAGCTGCTGGAGAAATGGAGG	1888
418	Qy	-----TyrGluGlnArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuLeuAs	434
1889	Db	TGTTGGCAACTGTGAACAAGTGATAAGAAAAACACAGAAAAATAATGGTCTTTCCTAGG	1948
434	Qy	nSerThrGlyAspLeuMetLysLysAsnTyrLeuHisAsnPheCysSerGlnValProSe	454
1949	Db	AGAAAAAGGCGCAATCCAACTGAAGA-----	1976
454	Qy	rValLeuGlyClnSerSerProLysValAlaSerLeuProSerIleSerVal-----	472
1977	Db	-GAGTTATCTCCCTGTCAGCCTGACATAAGGATGAATTCGAGTCCCAAGTATTAAGAAC	2035
473	Qy	----ProPheGlyGlyAlaArgProLysGlnProSerAsnLeuLysLeuGlnIleProLy	491
2036	Db	CCAAGCTGTTGGAGGGGCTCGACCTAAGCAG-----CTGCTGAGTCTTCCA--	2081
491	Qy	sProLeuSerAspHisLeuGlnAsnAspPheProAlaAsnSerGlyAsnAsnThrLysAs	511
2082	Db	-----CCAGGAACAGGAGTTCAAAGNACTGA--	2110
511	Qy	nLysAsnAspIleLeu-----GlyLysAlaLysLeuGlyGluAsnSerAlaTh	527
2111	Db	TAAGCCAGATGTTCTGGATGTCGCCAAGAGTGACCATGCACAGCAATGCTACTGTGT	2170
527	Qy	rAsnValCysSerProSerLeuGlyAsnIleSerAsnValAspThrAsnGlyGluHisLe	547
2171	Db	ATCCACATGCTCTCGACATCATATACCTGATCTTCAGGTAGCTTCAACTTCCAAATACAT	2230
547	Qy	uGluSerTyrGluAlaGluIleSerThrArgProCysLeuAlaLeuAla---ProAspSe	566
2231	Db	TGAC---ATAGAAAGTAATTTTGAAGATGGATCCAGTTTGTAACTCCAAACAAGGATTC	2287
566	Qy	z---ProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAlaArgLysProPheTh	585
2288	Db	TCTCCCTGAAAT-----	2317
585	Qy	rThrLeuGlyGluValAlaProValTrrValProAspSerGlnAlaProAsnCysMetLy	605
2318	Db	GGTTTTGGCCAGAAACAACTTACATGGGTCTTGATTCAGAGCTCCAACTGTATGAA	2377
605	Qy	sCysGluAlaArgPheThrPheThrLysArgArgHisHisCysArgAlaCysGlyLysVa	625
2378	Db	CTGCCAAGTCAAATTCACCTTTACAAAACGGCGACACCACTGCCGAGCATGTGGGAAGGT	2437
625	Qy	lPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTyrMetAspArgLysGluAl	645
2438	Db	ATTITGTGGTCTGTTGTAAACAGCAAAATGPAACCTTCAGTAICTAGG---AAGGAAGC	2494
645	Qy	aArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAlaTrpGluAsnMetMe	665
2495	Db	AAGAGTATGCTAATCTGCTATGAGACTATAAATAAAGCTCAGGCATTTGAAGAGATCAT	2554

Qy	665	tSerAlaSerSerGlnSerProAsnProAsnProAlaGluTyrCysSerThrIlePr	695
Db	2555	GAGTCCAGGTGGTCTTGTCTTAAGTCTAATCAATCTTAATGAA---TGTGCCCAT	2611
Qy	685	oProLeuGlnAlaGlnAlaSerGlyAlaLeuSerSerProProThrValMetVa	705
Db	2612	GCACATTCAGAGACCCAAACATCC-----AGTACACCTTCACCTACACTTGG--	2660
Qy	705	lProValGlyValLeuLysHisProGlyAlaGluValAlaGlnProArgGlnGlnArgAr	725
Db	2661	-CCCATCTCAGCACTGAAACAACCAATGTTGAAGGACCACTGTTCCAAAGAACAGAGAG	2719
Qy	725	gValTTPheAlaAspGlyIleLeuProAsnGlyGluValAlaAspAlaAlaLysLeu--	744
Db	2720	AGTATGTTTTCAGATGGCATCTGCCGATGTGTGAAGTTCGAGTACTACGAAATATTC	2779
Qy	745	-----	746
Db	2780	ATCTGGAAGTAAAGATGTTCTGATGACTTTTAGTCCNGTCTTACTCTGATGTGCCCAAT	2839
Qy	746	t-----AsnGlyThrSe	750
Db	2840	GATAAATAAAGTGGATCGTATCCCATCTCTCAGTGGAAAAACCAACAATGTTTAGG	2899
Qy	750	rSerAlaGlyThrLeuAlaValSerHisAspProVal-----	762
Db	2900	AGACATATTAAAGAGTAGATAGACCCAGAGTCCACACTTGTTCATACTGCACCTGTGGAAG	2959
Qy	762	-----	762
Db	2960	ACTGCCGGGAAACACAGGCACGGAAGCTTGGCCCATCGCTGGCCCTTTCACACTAGAAGA	3019
Qy	763	-----LysProValThrThrSerProLeuProAlaGluThrAs	775
Db	3020	TGATGTTTTTGTGGACAGCAGGAGCGCTCTACTCTACTGTGTGCTCTGCTAATAGTGG	3079
Qy	775	pIleCysLeuPheSerGlySer-----IleThrGlnValGlySerProValGl	791
Db	3080	TTTGCGTGTGTAGCATTTTCAGACTACAGGTATTGTGTAGTGTGCCAAATGTGTCTG	3139
Qy	791	ySerAlaMetAsnLeuIleProGluAsp-----GlyLeuProProlIleLeuIleSerTh	809
Db	3140	CAACAACATTAGTCTCTTACCACAGCATGACATGGTTGTGCTCCCTCTCGCCACATC	3199
Qy	809	rGlyValLysGlyAspTyr---AlaValGluGluLysProSerGlnIleSerValMetGl	828
Db	3200	TGGAGAGATGGATCAGTGCCTGTGTGGCAAGACGTCCCTCTCATGAGCAGATCAATTT	3259
Qy	828	nGlnLeuGluAspGlyGlyProAspProLeuValPheValLeuAsnAlaSerLeuSe	848
Db	3260	ACTTCTTTGAAGGGGAAGGATTTCCCTCGTGTACATTTGTCTAAATGCAACCTCTGT	3319
Qy	848	rMetValLysLleValAsnTyrValAenArgLysCysTyrCysPheThrThrLysGlyMe	868
Db	3320	GAATGTCAGTGTGTGCTTTATTCCTCAGAGAGTATGGTACTTCTCAACCAATGGATT	3379
Qy	868	tHisAlaValGlyIleGlnSerGluIleValIleLeuLeuGlnCysLeuProAspGluLysCy	888
Db	3380	GCATGGCTTGGACAGGCAGCAAAATTATGTCTCTGTTCCAAATGTTGCCAATGAGGATAC	3439
Qy	888	sLeuProLysAspIlePheAsnHisPheValGlnLeuTyrArgAspAlaLeuAlaGlyAs	908
Db	3440	GGTTCTTAAAGACATCTTCAGACTGTTTCATCACCATATATAGGATGCTTTAAAGAGAA	3499
Qy	908	nValValSerAsnLeuGlyHisSerPhePheSerGlnSerPheLeuGlySerLysGluHi	928
Db	3500	ATACATAGAAAAATTGGACAATCTGACCTTTTACTGACAGTCTTCTCAACGACGAAGATCA	3559
Qy	928	sGlyGlyPheLeuTyrValThrSerThrTyrGlnSerLeuGlnAspLeuValLeuProTh	948
Db	3560	TGGAGGGTTCCTGTTTATTACACTCTCCAGAACCTCTGATGGTCTCCGCTGGCCCAAG	3619

```

QY 948 rProProTyrLeuPheGlyLeuLeuIleGlnLysTrpGluThrProTrpAlaLysValPh 968
Db 3620 GAGTCCTTTCTCTCGGCATCTTATCCAGAGCTGGAGATTCCCTGGGCAAGAGTTT 3679
QY 968 eProfileArgLeuMetLeuArgLeuGlyAlaGluTyrArgLeuTyrProCysProLeuPh 988
Db 3680 TCCATATCGGCTAATGTTGCAATGGGAGCAGATACAAAGCATACCTGCTCTCTAAC 3739
QY 988 eSerValArgPheArgLysProLeuPheGlyGluThrGlyHisThrIleMetAsnLeuLe 1008
Db 3740 AAGTGTACAGCAGCAGCAACCTCTTTTGGAGAAATAGGACATATCTATTATGAATCTACT 3799
QY 1008 uAlaAspPheArgAsnTyrGlnTyrThrLeuProValValGlnGlyLeuValValAspMe 1028
Db 3800 TGTGATCTCGAATACCAAGTATACATTCGATACATTCGATCAGCTATTGATTCATAT 3859
QY 1028 tGluValArgLysThrSerIleLysIleProSerAsnArgTyrAsnGluMetLysAl 1048
Db 3860 GGAAATGGGAAGAGCTGTATAAATACCTCGAAAAGTACGATGATGATTAATGAAGT 3919
QY 1048 aMetAsnLysSerAsnGluHisValLeuAlaGlyGlyAlaCysPheAsnGluLysAlaAs 1068
Db 3920 AATACATTCCTTAACGAGCAGCTCATCAGTATTGGAGCTAGTTTATTAGTACAGAAGCAGA 3979
QY 1068 pSerHisLeuValCysValGlnAsnAspAspGlyAsnTyrGlnThrGlnAlaLysSerI 1088
Db 3980 TTTCATCTGCTGCTGTGACAGATGATGAGGATTTATCAGACTCAGGCCAACAGTGC 4036
QY 1088 eHisAsnGlnProArgLysValThrGlyAlaSerPheValPheSerGlyAlaLeuLys 1108
Db 4037 CACTGGTCAGCAAGCAAGTGCAGAGTTCGCAAGTTTGTGGTGTATGATGGTCTGAA 4096
QY 1108 sSerSerSerGlyTyrLeuAlaLysSerSerIleValGluAspGlyValMetValGlnI 1128
Db 4097 AACATCTTCAGGATTTCTTCGCAATCCAGCATAGTCCGAATGCTTAATGTGACAAAT 4156
QY 1128 eThrAlaGluAsnMetAspSerLeuArgGlnAlaLeuArgGluMetLysAspPheThrI 1148
Db 4157 AACTCCAGAGACCATGAGGCTTGGCGCTGCTTTACGAGAGCAAGACATTTAGAA 4216
QY 1148 eThrCysGlyLysAlaAspAlaGluGluProGlnGluHisIleHisIleGlnTrpValAs 1168
Db 4217 CCAGTGTGGAAGATGATGCGGTAGACCTGAGAGAGATGTTGGACATCTGTTGGGTGA 4276
QY 1168 pAspAspLysAsnValSerLysGlyValValSerProIleAspGlyLysSerMetGluTh 1188
Db 4277 TTCAGAGAGAGGAGCAAGAGTGTATCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 4336
QY 1188 rIleThrAsnValLysIlePheHisGlySerGluTyrLysAlaAsnGlyLysValIleAr 1208
Db 4337 ATTTCCAAAGTGAATAAATAAACTGGAAACAGATTTTGAACCTCAGCAGAGAGCTGTGA 4396
QY 1208 gTrpThrGluValPhePheLeuGluAsnAspGlnHisAsnCysLeuSerAspProAl 1228
Db 4397 ATGCACCTGAGGTGTCTACTCTCTAAAGGACCAAGATATATCTATTATCA----- 4448
QY 1228 aAspHisSerArgLeuThrGluHisValAlaLysAlaPheCysLeuAlaLeuCysThrGl 1248
Db 4449 -AGTAGTACCAAGTTCCTCAAGAGGATGCTGTGGCTGTGTAGTCTGCCCTGTGCCCTCA 4507
QY 1248 nLeuLysLeuLeuLysGlyAspGlyMetThrLysLeuGlyLeuArgValThrLeuAspSe 1268
Db 4508 CCTGAGAACTCTAAAGAGTAACAGATGAACAAATTTGGAATGCACTCAGAGTGTCCATTCAC 4567
QY 1268 rAspGlnValGlyTyrGlnAlaGlySerAsnGlyGlnHisLeuProSerGlnTyrMetAs 1288
Db 4568 TGATATGTTGAGTTCAGCAGCAGCTGTGAAGCCAGCTTCTGCCCTCAGCATTTACCTGAA 4627
QY 1288 nAspPheAspSerAspLeuLysMetIleHisGlyGlyAlaCysGlnLeuSerGluGl 1308
Db 4628 TGATCTGATAGTCCCTGATCTCCATCCAGCGGAGCTCCCAAC---TCTAGTCT 4684
QY 1308 yProValValMetGluLeuIlePheTyrIleLeuGluAsnIle 1322

```

```

Db 4685 GCCACTAGAAATAGAGTTAGCATTTTTCATCTTAGAAAACCTT 4727
RESULT 14
AKI22241
LOCUS Mus musculus mRNA for mKIAA0305 protein. linear ROD 15-MAR-2003
DEFINITION AKI22241.1 GI:28972138
ACCESSION AKI22241
VERSION AKI22241.1
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Okazaki,N., Kikuno,R., Ohara,T., Inamoto,S., Aizawa,H., Yuasa,S.,
Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA
gene: II. The complete nucleotide sequences of 400 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences
of cDNA clones randomly sampled from size-fractionated libraries
JOURNAL DNA Res. 10, 35-48 (2003)
AUTHORS Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics, 2-6-7
Kazusa-kamatori, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.
FEATURES
Location/Qualifiers
1..5438
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbg08215"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"
1..5438
/gene="mKIAA0305"
/locus_tag="mKIAA0305"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence=not experimental
/product="mKIAA0305 protein"
/db_xref="GI:28972138"
/translat="VLSKPTGRMDSYFKAASVGLDQLDDPEQNPDKQDYLDPDAVAFN
QCVSSSEASPOLALLSKDQRCISTCASSEACEDANETFLGKTHEGLTSRPNKKN
AGLDLLSVDASTDSEIOPCSMRSCSPVCDLISDMGNLHATNSEEIYKOLLPPDK
SSADTLITLDSYSEALTVSVCDSNAVREEQNNINAGINRDISIKELGVKVDMA
LDGCKNRTNENKDLKILSNELETVDPMPSVLMESQSEMENTKDNPOYKPLFCCLL
FLQENRTSFMKQTVKEDRNLDKNNNDIVHVSDDVPPSLSCLSLSSGLSGSLIHN
NEHSDILPPNESSEGNDAVTHBEIQKSDVLGDTDLSKETCRSIFLOPVNEKKGE
KQVEEVMVIGESLESPEASAAAGSPVALSAASVPEAGPCEGTFPSSDDMGQ
ELDYFNIDESMRSGLISDAELDAFLKQCILSNSTMSAGENVDSOLQMNQITMKG
HDEVAGDIYFNAEAGAGENGCGVNCETSDKENTENNGLSIGEKGAITRELSACQP
DIRDLPLPDSIKTQVAGRPFKLLSPFTSRSKLNKPDVDPVPESEPTANATAV
STCSADHLPDSQVFNISYDIENFEDSGSFVTANKDSLLENKRKESVLGQKQPTW
VPDEAPNMCNQCVRFTTKRRHRCAGKFCVCCNCKQLQYLEKEARVCVYE
TINKAQAERMWSPGSGCLKSHNECATDQLOETQTSPTSPSTLPIISALKQPNVE
GPCSKQKRVFADGILPNEGAVDTKLSSGKSCSDDFSVLPVPTPMINKVDRTHS
PTVEKPNGLGDIIRSETSQSPCTHTAPVELPGNTGEGIPMPCPTLEDVDFVDS
EPSTPTVPASGILPVASISYRLICSAKVCVNNISLLPDDDLGLPLLATSDGDS
PVQVQRPSEHQIILLGEGFPFATFVNANLVNKLVSSEKYSKYSYFSTNGLHGL
GQAEIIVLLQCLPNEDEVPKDIFRLFITIYKDALKGKIENLDNLTFTEFLNSKDHG

```



GFLPITFTFQNLQDLVPRSPFFLCGILIQKLEIPWAKVPPMRLMLBLGAEYKAYPAPL  
 TSVRKLPLFGEIGHTIMNLLVDRNYQYTLNIDQLLHMEMGKSCIKIPRKXSDV  
 MKVHNSNEHVIIGASFTSDADSHLVQSDGVYQOANSATGQPRKVTGASFFVFN  
 GALKTSFGFLAKSIVSDGLMVLQITPTMETEGLRLALREQKDFRQCGKVDVLDREYV  
 DICWDSEERKNKGVISVDGMVGEPSPSEKIKLETFDETEKTVKCTEYVFLKDDQ  
 ISILSSYQPAKEIAVACSAALCPHLRTLAKSNRMKIGLRVSDTDWVFOAGCEQOL  
 LPQHYLNDLSALIPVHGTSNSSUPLLELFAFILENLSE"

ORIGIN

Alignment Scores: 5.19e-126 Length: 5438  
 Pred. No.: 2045,00 Matches: 523  
 Score: 2045,00 Matches: 523  
 Percent Similarity: 49.56% Conservative: 208  
 Best Local Similarity: 35.46% Mismatches: 481  
 Query Match: 29.26% Indels: 264  
 DB: 10 Gaps: 40

US-09-744-167-2 (1-1323) x AK122241 (1-5438)

QY 3 IleAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrSerTrpAsp 22  
 Db 613 TTGACGTCGTCCTCAGTTCAGAG----- 636  
 QY 23 GlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeuProAsp 42  
 Db 637 -----GCTCTCAGTCTCTCTCAGTACGTCGTAGTATGCTGCAGAGAAGAG 687  
 QY 43 GluIysAsnValLeuValAlaValMetHisAsnCysAspLysArgThrLeuGlnAsn 62  
 Db 688 CAGATAACATC-----AACGCTGTGTATAAAACACAGATATT 726  
 QY 63 AspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSerCysSer 82  
 Db 727 AGTCAAGAAGATGGGTGTAAGAGTAAAGTACATGCTTTTGTAT-----TCTTTGAAA 780  
 QY 83 LeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSerThrGlu 102  
 Db 781 TATAATAGACAGAAAATTTAAGGATATAAATTTCTTAATGAGTTAGAAACAGTTGAT 840  
 QY 103 LysAspMetAsnSer-----GluLysGlnMetAspProLeuAsnArgProLysThrGlu 120  
 Db 841 TTTGACATGCCATCTGTTTGTATGAAACAAAGTCTGTAAGTCTAATACCAAGATAAC 900  
 QY 121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140  
 Db 901 CCACAAACAGAGGCTGCCATG----- 924  
 QY 141 ProSerGlnLeuLysAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160  
 Db 925 ---GAATTGTTAAAGATGATGGCTGTTTGGCAGAGAGAAAGTAGCTGGCGAGTCAAT 981  
 QY 161 -----SerLeuThrValAsp-----SerVal 167  
 Db 982 AATACAGAGTGTTTAGAAAGAGGAGGAGGAAGTAACACAAATGCTATGCTTGCAGAGCTT 1041  
 QY 168 IleSerSerGlnGlyThrAspGlyCysProAlaValLysLysGlnLeuAsnTyr---Ile 186  
 Db 1042 CCAGAAAATGAGGAGATATCCCAAGTACCCAGCTTCAAAAGATGAAATTTCAAGTTA 1101  
 QY 187 ProAspGluAspLeu-----ThrGlyLysIle 195  
 Db 1102 CTGACTTCCCTCCAGGAAAATAGAACTTCTGTGTTTCATGAAACAGACTGTAAAGAA 1161  
 QY 196 SerSerProArgThrAspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIle 215  
 Db 1162 GACTCAAGAAAATTTAGACCTTAAAGATAATAATGATATAGTCCATGTTTCA----- 1212  
 QY 216 LeuMetLysLysGluProAlaGluSerThrThrGluGluSerLeuArgSerGlyLeu 235  
 Db 1213 -----GGTATGATGTGCCCTTCACTGCTCTCTCTTTCATTTCTGCTGGGTCTTTG 1263  
 QY 236 ProLeuLeuLysProAspMetProAsnGlySerGlyArgAsnAspCysGluArg 255

Db 1264 -----TGTGGATGTTTAATTTCAATAAT-----GAA 1290  
 QY 256 CysSerAspCysLeuValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHis 275  
 Db 1291 CACAGTGAATTTTACCTCCCTAAATGAGTCTGAAGCCAGAAATGATGAGTGAATATA 1350  
 QY 276 GluGluThrLeuGlyThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGln 295  
 Db 1351 CATGAAGAGATACAGACAGAGTGTCTTTTACATGGGAGACAGATCTTTCAAGAGAGAA 1410  
 QY 296 AspMetThrAsnTrpLysLeuThrLysLeuAsnGlnMetAsnAspSerGlnValAsnGlu 315  
 Db 1411 ACGTGTAGGAGCATATTTTCCAGCAGTAATAAGAAAGAGGGAGGAGGAGTTGAG 1470  
 QY 316 GluLysGluLysPheLeu-----GlnIleSerGlnProGluAspThrAsnGly--- 331  
 Db 1471 GTTGAGGAGATGGTAATCAGTGGTGAAGTCTTTGGAGTCCCTGAGGAGCCAGCTCTGT 1530  
 QY 332 -----AspSerGlyGlyGln 336  
 Db 1531 GCAGCAGCTGGGTCTCCGGTGGCACTTTCTGTCGCAAGTGTCCAGAGGCTCTGTGTC 1590  
 QY 337 CysValGlyLeuAlaAspAlaGlyLeuAspLysGlyThrCysIleSerGluSerGlu 356  
 Db 1591 TGTGAAGTCTCCTTTCCAGCAGTATATGATGGTGG-----CAA 1632  
 QY 357 GluCysAspPheSerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCys 376  
 Db 1633 GAGTTAGATTACTTTAATATGATGAAGCATGAGACATGAGCATTAATTAATGAGCA-TGC 1691  
 QY 377 AspSerTyrGlyMetGlnAspProGlyValSerPheVal-ProLysThrLeuProSerLys 396  
 Db 1692 TGAACCTCGATGTTTTCGAAGAAACAGTGTCTTTGAACTCTACACAAATGTTCTGTGG 1751  
 QY 396 sglu-----AspSerValThrGluGluLysGluIle-----G1 407  
 Db 1752 AGAAATGTAATGATTCAGTTCAGATGAATCAGATCACTATGAAAGGACTACATGA 1811  
 QY 407 uGluSerLysSerGluCysTyrSerAsnIle----- 417  
 Db 1812 TGAATAATGTCGGGATATATCTTCAATCTGTAAGCAGAGAGTGTCTGGAGAAATGAGG 1871  
 QY 418 -----TyrGluGlnArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuLeuAs 434  
 Db 1872 TGTGGCAACTGTGAACCAAGTATAAGAAACACAGAAATAATGTTCTTCCATAGG 1931  
 QY 434 nSerThrGlyAspLeuMetLysAsnTyrLeuHisAsnPheCysSerGlnValProSe 454  
 Db 1932 AGAAAAGGCGCAATTCACACTGAAAGA----- 1959  
 QY 454 rValLeuGlyGlnSerSerProLysValValAlaSerLeuProSerIleSerVal----- 472  
 Db 1960 -GAGTATCTGCTGTGCTGACCTGACATAGGATGAATGCGAGTCCCGCAGTATTAAC 2018  
 QY 473 ----ProPheGlyAlaArgProLysGlnProSerAsnLeuLysLeuGlnIleProly 491  
 Db 2019 CCAAGCTGTGGAGGGCTCGACCTAAGCAG-----CTGCTGAGTCTTCCA-- 2064  
 QY 491 sProLeuSerAspHisLeuGlnAsnAspPheProAlaAsnSerGlyAsnAsnThrLysAs 511  
 Db 2065 -----CCAGAACAGGAGTTCAAGGAAGTCA 2093  
 QY 511 nLysAsnAspIleLeu-----GlyLysAlaLysLeuGlyGluAsnSerAlaTh 527  
 Db 2094 TAAGCCAGATGTGTGATGTCAGAAAGTGCAGCATGCACAGCAAAATGCTACTGCTGT 2153  
 QY 527 rAsnValCysSerProSerLeuGlyAsnIleSerAsnValAspThrAsnGlyGluHisLe 547  
 Db 2154 ATCCACTGCTCTGCAGATCATATACCTGATTTCTCAGGTAGCTTCACTCACTCAATATCAT 2213  
 QY 547 uGluSerTyrGluAlaGluIleSerThrArgProCysLeuAlaLeuAla-----ProAspSe 566  
 Db 2214 TGAC---ATAGAAAGTAAATTTTCAAGATGATCCAGTTTGTAACTGCAACAGAGGATTC 2270



```

Db      4380  ATGCACCTGAGGTGTTCTTACCTCTCTAAAGGACCAAGATATATCTATTTATCA----- 4431
Qy      1228  aAsHisSerArgLeuThrGluHisValAlaLysAlaPheCysLeuAlaLeuCysThrGlu 1248
Db      4432  -AGTAGTTACCAAGTTTCCAAAGGAGATGCTGTGGCTTGAGTGCCTCGCCCTCA 4490
Qy      1248  nLeuLysLeuLeuLysGlyAspGlyMetThrLysLeuLysLeuArgValThrLeuAspSe 1268
Db      4491  CCTGAGACTTAAGAGTACAGAAAGACAAATGGACTCAGAGTGCCTTGCAC 4550
Qy      1268  rAspGlnValGlyTyrGlnAlaGlySerAsnGlyGlnHisLeuProSerGlnTyrMetAs 1288
Db      4551  TGATATGTTGAGTTCAGGAGGAGCTGTGAAGGACAGCTTCTGCCTCAGCATTACCTGAA 4610
Qy      1288  nAspPheAspSerAspLeuValLysMetIleHisGlyGlyAlaCysGlnLeuSerGluGlu 1308
Db      4611  TGAATCTGATAGTCCCTGATCCCTGATCCAGCCAGCGGACCTCCAC---TCTAGTCT 4667

1308 yProValValMetGluLeuIlePheTyrIleLeuGluAsnIle 1322
Db      4668  GCCACTAGAAATACAGTATGACATTTTTCATCTAGAAAACCTT 4710

RESULT 15
HSM805726
LOCUS   Homo sapiens mRNA; cDNA DXFZp686K236 (from clone DXFZp686K236);
        complete cds.
ACCESSION BX537424
VERSION   BX537424.1 GI:31873349
KEYWORDS Homo sapiens (human)
SOURCE   Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 6280)
AUTHORS  Ansorge, W., Krieger, S., Regiert, T., Rittmüller, C., Schwager, B.,
        Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and
        Wiemann, S.
        Direct Submission
        Submitted (16-JUN-2003) MIPS, Ingolstaedter Landstr.1, D-85764
        Neuherberg, GERMANY
        Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
        Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
        sequenced by EMBL (European Molecular Biology Laboratories,
        Heidelberg/Germany) within the cDNA sequencing consortium of the
        German Genome Project.
        This clone (DKFZp686K236) is available at the RZPD in Berlin.
        Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
        Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
        information about the clone and the sequencing project is available
        at http://maps.ssf.de/proj/cDNA/.
        Location/Qualifiers
            1..6280
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /map="5q14.1"
               /clone="DKFZp686K236"
               /tissue_type="human cervix"
               /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
               DH10B; sites SfiII + SfiIB"
               /dev_stage="adult"
            1..6280
               /gene="DKFZp686K236"
               /3'..4752
               /gene="DKFZp686K236"
               /note="endofin"
               /codon_start=1
               /product="hypothetical protein"
               /protein_id="CAD97666.1"
               /db_xref="GI:31873350"
               /translation="MDSYFKAAYSDLDKLLDFEQNPDEQDYLQVQNAVDSNHQSVS

```

```

SELASSORTSLLPKQPCVNSCASSSETSYGTNESSLNKTLKGLTSTQNKENVLTGLDL
LSSVDGTSDEIQPLMGRSKPICDLIISDMGNLHATNGEEDIKLLPDPKSNADS
LIGLDLSVSTPCVSTSDHSDTVRQNDTSELQNRBIGGIKEIGIKVDTITLSDS
LYNVESTNLKDKIFNOLIESIVDFNMSSALTROSSKMFHAKDKLQHSQPCGLLKVDG
LVKEEVAVALTAECLEKSGKTSALTSALPKQEDLCNDNSRDNFKLPDPFSFOD
KTVLQKQABEDSKLDLKDNDVIQSSSALHVSCKDVPSSLSCLPASGMSGSLIESK
ARGDFLOHEKONIQAATIHBEIQNSVVLGEPFKENDLLQKCKSKILLQSLLEG
MEDRKIDPQTVIRAESLDGDSSTVVEGSLGTHVPESSDCCGFIINTFSNDM
DQGLDYNFDIGAGSLGIESDAELDAFTQYLQTTNIKSFEENVDSKSNQNDIM
KGLDGNINITYFNAEAGAISHGINII CETVDKONTIENGLSAGSKSTIPVQOGLP
TSKEITNOLSDINSQSGVARGPKOLFSLPSRTRSSKDLNKPDPDTTESPSTAD
TWPTICNADSDAPQVSNFNSYDIIESNSEGSSFTANEDSVPENTCEGLVGLQK
QTPWPSAEAFNMNQCQKFTTKRHHCRACGKFCGVCNCRKCKLOYLEKBAVCV
VCYITISKAQAFERMTSGTSLKNSHDECTIVQPPQENQTSIPSPFALPVSAALQK
PGVGLCSKEQRVMFADGILPNEGVEDATITKLSGSKRCSBEDFSPSPFPMVTMVD
HSHETTVKPNNETGDIRNEIIQSIQSPVSEKLSMTNCEGLPTSGSTLDDDFV
AETPEPSPTGLVNSNLPIASISDYELLCDINKVCNKISLLPNDDSDSLPPLVAGS
EKGSPVVEHPHSEQIILLEGEPHPTFVLNANLVNKKIFYSDDKYWYFSTNG
LHGLQABEIIILLCLPNEPTPKDIPELITLYKDALKGKLTENLONIPTSFSLSS
KDHGFLITITFORLDDLSLPSNPFGLGILIQLEIPWAFPMRLMLBLGAEYKAY
PAPLTSIRGRKPLFEIGHTIMNLLVDNRNYQTLHNDIQLLHWMGKSKCIKPRKK
YSDVMKLVNSNEHVISIGASFSTEADSHLVCIQNDGITYQANSTTGHPRKVTGASF
VVFNGALKTSGFLAKSIVEDGLMVOITPTETMNGRLALREQKDFKITCGKVDVGL
REYVDICWDAEERKNGKIVSSVDGISIQGPSPSEKILEADFETDEXIKVCTEYFVL
KDQDLSILSTSYQPAKEIAMACSAALCPHLKTLKSNMKNIGLRVSDTDWVEFQAGS
EQGLLPQHLNLDLSALIPVHGTSTNSSLPLELELVFFIIEHLF"

ORIGIN
Alignment Scores:
Pred. No.: 1,01e-125 Length: 6280
Score: 2042.00 Matches: 541
Percent Similarity: 50.03% Conservative: 222
Best Local Similarity: 35.48% Mismatches: 476
Query Match: 29.21% Indels: 286
DB: 9 Gaps: 51

US-09-744-167-2 (1-1323) x HSM805726 (1-6280)
Qy      4 AspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrSerTyrAspGln 23
Db      412  GAAAAAATGTAACAGGACTTGATCTCTTCTCTCTGATGTTGATGTTGATGAA 471
Qy      24  CysSerAlaValGluValGlyLysCysGlyAsnLeuAlaCys-----LeuPro 41
Db      472  ATCCAGCGCTTATATATCGGA-----CGATGTAGTAAACCTATCTGTGATCTGATAAGT 525
Qy      42  AspGluLysAsnValLeuValAlaValMetHisAsnCysAspLysArgThrLeuGln 61
Db      526  GACATGGGTAACTAGTTCATGCAACCAATAGTAGAAGATATATAAAATTTATGCCA 585
Qy      62  AsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeu-----MetAspAlaPhe 79
Db      586  GATGATTTTAACTCT-----AATCAGATCTCTTGTGATGATGATTTATCT 633
Qy      80  SerCysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGlu 99
Db      634  TCAGTGTGATGATCTCCCTCTGTTTCTTCAACAGACCATGATGATGATCTGTCAGAAA 693
Qy      100 SerThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThr 119
Db      694  CAA---CAGATGATACAGTCTTCAATTAACAAATAGAGAAATCGGAGGAATCAAGAA 750
Qy      120 GluGlyArgSerValAsnHisLeuCysProThrSerSerAspSer----- 134
Db      751  TTGGGTATAAAAGTAGAT-----ACAACACTTTTCAGATCTCTATAATTTACAGTGA 801
Qy      135  -----LeuAlaSerValCysSer----- 140
Db      802  ACAGAAATTTAAAGATAAAAGATCTTTAATCAGTTAGTAATCAATTTGATTTTAAAC 861
Qy      140  ----- 140
Db      862  ATGTCTATCTGCTTTTGACTCGCAAAAGTTTCCAAATGTTTTCATGCCAAAGACAGCTACAA 921

```

QY 141 ---ProSerGln-----LeuIysAspAspGlySerIleGlyArgAspProSer 155  
Db 922 CACAGAGCCCGCATGGGATTACTAAAGATCTTGGCTAGTAAAGAGGAGTAGAT 981  
QY 156 MetSerAlaIleThrSerLeuThrValAspSerValIleSerSerGlnGlyThrAspGly 175  
Db 982 GTGGCAGTCATACTGCGCA-----GAATGTTTAAAGAGAGGCAAGACAAGT 1032  
QY 176 CysProAlaValLysLysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysIle 195  
Db 1033 GCTTTGACCTGCAGCCTTCCGAAAAT-----GAAGATTATGCTTAAATGAT 1080  
QY 196 SerSerProArgThrAspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIle 215  
Db 1081 TCARATTCAGAGATGAATTCCTCAATTAATTAATTAATTAATTAATTAATTAAT 1140  
QY 216 LeuMetLysLysGluProAlaGluGluSerThrThr----- 227  
Db 1141 GTTATAAAACAATCTGCACAGAGACTCAAAAAGTTTAGACCTTAAAGGATAATGATGA 1200  
QY 228 GluGluSerLeuArgSerGlyLeuProLeuLeuLeuLysProAspMetProAsnGlySer 247  
Db 1201 ATCCAGATTCCTCTCAGCTTTACATGTTTCCAGTAAA---GATGCGCGTCTCTCATTTG 1257  
QY 248 GlyArgAsnAsnAspCysGluArgCysSerAspCysLeuValProAsnGluValArgAla 267  
Db 1258 TCCGTCTCTCTGGCTCTGGGTCTATGAGTGGATCATTAATGAAGAGTAAGACCGGGT 1317  
QY 268 AspGluAsnGluGlyTyrGluHisGluGluThrLeuGlyThr----- 282  
Db 1318 GATTTTTCCTCAGCATGAACATAAAGATAATATACAGATGCAGTGACTATACATGAA 1377  
QY 283 GluPheLeuAsnMetThr-----GluHisPheSerGluSerGlnAspMetThr 298  
Db 1378 GAAATACAGACAGCTGTGTTCTAGTGGGGAACCATTCAGAGAGATGATCTTTTGAAA 1437  
QY 299 AsnTrpLys-----LeuThrLysLeuAsnGlu---MetAsnAspSerGln 312  
Db 1438 CAGGAAAAATGTAAAGACATCTCTTCAGTCATTAATTAAGGGATGGAAGACAGAAAG 1497  
QY 313 ValAsnGluGluLysGluLysPheLeuGlnIleSerGlnProGluAspThrAsnGlyAsp 332  
Db 1498 ATAGCTCTGACCCAG-----ACAGTAAATCAGAGCTGAGTCTTTGGATGGTGT 1545  
QY 333 SerGlyGlyGlnCysValGlyLeuAlaAspAlaGlyLeuAspLeuLysGlyThrCysIle 352  
Db 1546 GACACCAGTCTACAGTGTAGATCTCAAGAGGG-----CTTTCTGGCACTCATGTC 1599  
QY 353 SerGluSerGluGluCysAspPheSerThrValIleAspThrProAlaAlaAsnTyrLeu 372  
Db 1600 CCAGAGTCTTCTGATGTG---TGGAAGGTTTTTATTATTAATCTTTTCAAGCAATGATG 1656  
QY 373 SerAsnGlyCysAspSerTyrGlyMetGlnAspProGlyValSerPheValProLysThr 392  
Db 1657 GATGGGCAAGACTTAGATTACTTTAATTATGATGAGCGCAAAAGTGCCCACTAATT 1716  
QY 393 LeuProSerLysGluAspSer---ValThrGlu----- 402  
Db 1717 AGTGATGCTGAACCTGATGCTTCTGCAGAGACAGTATCTTCAGACCACTAATAAAG 1776  
QY 403 -----GluLysGluIleGluGluSerLysSerGluCys----- 413  
Db 1777 TCITTTGAAGAAATGAATGACTCTAATTCGGAATGAATGAATCAGATAGATATGAAGGC 1836  
QY 414 -----TyrSerAsnIleTyr---GluGlnArgGlyAsnGluAlaThrGlu 427  
Db 1837 TTAGATGATGGAACATCAATAATATATTTTCAATCAGAGCAGGAGCTATTGGGAA 1896  
QY 428 GlySerGlyLeu---LeuLeuAsnSerThrGlyAspLeuMetLysLysAsnTyrLeuHis 446  
Db 1897 AGTCATGGTATTATATAATTTGTGAACACAGTTGAT-----AAACAAAATACAAATAGAA 1950

QY 447 AsnPheCys-----SerGlnValProSerValLeuGlyGlnSerSerPro 461  
Db 1951 AATGGCCTTCTTTTAGGAAAAAAGCACTATTTCAGTTTCAAAAGGGTTACCTACCACT 2010  
QY 462 Lys-----ValValAlaSerLeuProSerIleSerValProPhe 474  
Db 2011 AAGTCTGAGATTACAATCAATATATCATCTCTGTATATTAAACAGTCAATCTGTT----- 2064  
QY 475 GlyGlyAlaArgProLysGlnProSerAsnLeuLysLeuGlnIleProLysProLeuSer 494  
Db 2065 GGAGGGGCCAGACCTAAGCAATTGTTTAGCCTT-----CCATCAAGA 2106  
QY 495 AspHisLeuGlnAsnAspPheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAsp 514  
Db 2107 ACAAGGAGTTCAAGACCTG-----AATAAGCCAGAT 2139  
QY 515 IleLeuGlyLysAlaLysLeuGlyGluAsnSerAlaThrAsnVal----- 529  
Db 2140 GTTCCAGATACAAATAGAAAGTGAACCCAGCAGCAGATACCGTTGTTTCCAATCACTTGT 2199  
QY 530 -----CysSerProSerLeu-----GlyAsnIleSerAsnValAsp 541  
Db 2200 GCTATAGATCTACAGCTGATCCAGGTTAGCTTCAACTCTAATTAATTAATGATTAGAA 2259  
QY 542 ThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeuAla 561  
Db 2260 AGTAATTTCTGAAGTGGATCTAGTTTC-----GTAAC 2292  
QY 562 LeuAlaProAspSer---ProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580  
Db 2293 GCAATAGAGATTTCTGTACCTGAAACACTTGCAAAGAGAGC----- 2334  
QY 581 ArgLysProPheThrThrLeuGlyGluValAlaProValIleValProAspSerGlnAla 600  
Db 2335 -----TTGGTTTGGCCAGAAACAGCCCTACTTGGTTCTCTGATTCAGAGCT 2382  
QY 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg 620  
Db 2383 CCAAACTGTATGAATGCGCAAGTCAATTTACTTTTACAAACGCGCAGACCATTCGCGA 2442  
QY 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLysMet 640  
Db 2443 GCATGTGGGAAAGTATTTTGTGGTCTGTGTGTATAGGAAGTGTAAACTGCAATATCTTA 2502  
QY 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAla 660  
Db 2503 GAA---AAGGAAGCAAGAGTATGTCTAGTCTGTATGAAACTATTAGTAAAGCTCAGGCA 2559  
QY 661 TrpGluAsnMetMetSerAlaSerSerGlnSerProAsnProAsnAsnProAlaGluTyr 680  
Db 2560 TTTGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2616  
QY 681 CysSerThrIleProLeuGlnGlnAlaGlnAlaSerGlyAlaLeuSerSerProPro 700  
Db 2617 TGTACTACTGTCCAGCTCTCAGAGAACCAACATCC-----AGTATACCTTCA 2667  
QY 701 ProThrValMetValProValGlyValLeuLysHisProGlyAlaGluValAlaGlnPro 720  
Db 2668 CCAGCAACTTTG---CCAGTCTCAGCACTTAAACCAACCAAGGCTGTGGAAGGACTATGTTCC 2724  
QY 721 AsgGluGlnArgValTrpPheAlaAspGlyIleLeuProAsnGlyGluValAlaAsp 740  
Db 2725 AAAGAACAGAGAGATGTTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2784  
QY 741 AlaAlaLysLeuThrMetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAsp 760  
Db 2785 ACAACAAAATTA-----TCATCTGGAAGTAAAGAGATGTTCTGAAGAC 2826  
QY 761 -----ProValLysProValThrThr----- 767  
Db 2827 TTTAGTCTCTCTCAGCTGATGTGCTATGACAGTAAACACAGTGGATCATTTCCCATCTCT 2886  
QY 767 ----- 767

